LETTERS

Protein structure determination in living cells by in-cell NMR spectroscopy

Daisuke Sakakibara[1,2]*, Atsuko Sasaki[1,2]*, Teppei Ikeya[1,3]*, Junpei Hamatsu[1,2], Tomomi Hanashima[1], Masaki Mishima[1,2], Masatoshi Yoshimasu[4], Nobuhiro Hayashi[5]†, Tsutomu Mikawa[6], Markus Wälchli[7], Brian O. Smith[8], Masahiro Shirakawa[2,9], Peter Güntert[1,3,10] & Yutaka Ito[1,2,6]

Investigating proteins 'at work' in a living environment at atomic resolution is a major goal of molecular biology, which has not been achieved even though methods for the three-dimensional (3D) structure determination of purified proteins in single crystals or in solution are widely used. Recent developments in NMR hardware and methodology have enabled the measurement of high-resolution heteronuclear multi-dimensional NMR spectra of macromolecules in living cells (in-cell NMR)[1–5]. Various intracellular events such as conformational changes, dynamics and binding events have been investigated by this method. However, the low sensitivity and the short lifetime of the samples have so far prevented the acquisition of sufficient structural information to determine protein structures by in-cell NMR. Here we show the first, to our knowledge, 3D protein structure calculated exclusively on the basis of information obtained in living cells. The structure of the putative heavy-metal binding protein TTHA1718 from *Thermus thermophilus* HB8 overexpressed in *Escherichia coli* cells was solved by in-cell NMR. Rapid measurement of the 3D NMR spectra by nonlinear sampling of the indirectly acquired dimensions was used to overcome problems caused by the instability and low sensitivity of living *E. coli* samples. Almost all of the expected backbone NMR resonances and most of the side-chain NMR resonances were observed and assigned, enabling high quality (0.96 ångström backbone root mean squared deviation) structures to be calculated that are very similar to the *in vitro* structure of TTHA1718 determined independently. The in-cell NMR approach can thus provide accurate high-resolution structures of proteins in living environments.

In living cells, proteins function in an environment in which they interact specifically with other proteins, nucleic acids, co-factors and ligands, and are subject to extreme molecular crowding[6] that makes the cellular environment difficult to replicate *in vitro*. Although *in vitro* methods of structure determination have made very valuable contributions to understanding the functions of many proteins, *in vivo* observations of 3D structures, structural changes, dynamics or interactions of proteins are required for the explicit understanding of the structural basis of their functions inside cells. Its non-invasive character and ability to provide data at atomic resolution make NMR spectroscopy ideally suited for the task. Indeed, recent advances in measurement sensitivity have permitted heteronuclear multi-dimensional NMR spectroscopy of proteins inside living cells by so-called in-cell NMR[1–5].

In-cell NMR has been used to detect protein–protein interactions inside *E. coli* cells[7] as well as the behaviour of intrinsically disordered proteins[8,9]. In eukaryotic cells, in-cell NMR studies have been performed by injecting proteins into *Xenopus laevis* oocytes or eggs[10,11] and, more recently, cell-penetrating peptides have been used to deliver proteins that can be observed in living human cells[12].

Despite the interest in in-cell NMR, it has not yet been determined whether the established methods for the structure determination of purified proteins by NMR[13] can be extended to proteins in living cells. To our knowledge, the only published 3D in-cell NMR experiments

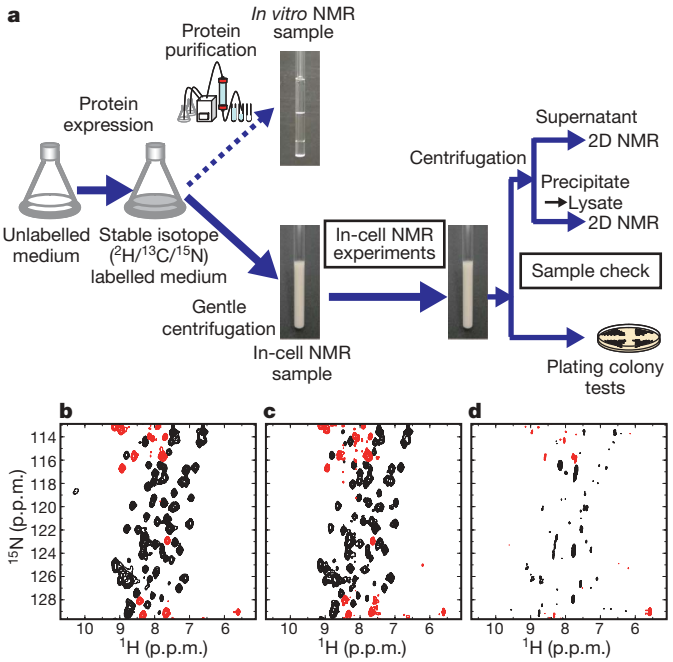

Figure 1 | Stability of *E. coli* cells expressing TTHA1718 under NMR measurement conditions. **a**, Scheme of the in-cell NMR experiments using *E. coli* cells. **b**, The ^{1}H–^{15}N HSQC spectrum of a TTHA1718 in-cell NMR sample immediately after sample preparation. **c**, The ^{1}H–^{15}N HSQC spectrum after 6 h in an NMR tube at 37 °C. **d**, The ^{1}H–^{15}N HSQC spectrum of the supernatant of the in-cell sample used in **b** and **c**.

[1]Department of Chemistry, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan. [2]CREST/Japan Science and Technology Agency (JST), 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. [3]Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance, J. W. Goethe-University Frankfurt, Max-von-Laue Straße 9, 60438 Frankfurt am Main, Germany. [4]Cellular and Molecular Biology Laboratory, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan. [5]Division of Biomedical Polymer Science, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake-shi, Aichi 470-1192, Japan. [6]Research Group for Bio-supramolecular Structure-Function, RIKEN, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. [7]Bruker BioSpin, 3-21-5 Ninomiya, Tsukuba-shi, Ibaraki 305-0051, Japan. [8]Division of Molecular and Cellular Biology, Faculty of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK. [9]Department of Molecular Engineering, Graduate School of Engineering, Kyoto University, Nishikyo-Ku, Kyoto 615-8510, Japan. [10]Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany. †Present address: Department of Life Science, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 B-1, Nagatsuda-chou, Midori-ku, Yokohama, Kanagawa, 226-8501, Japan. *These authors contributed equally to this work.

are projection reconstruction versions of triple-resonance experiments for backbone resonance assignment in *E. coli*[14]. *De novo* NMR protein structure determination in living cells requires methods for resonance assignment that do not rely on information obtained *in vitro*, as well as the efficient analysis of nuclear Overhauser effect (NOE)-derived information, in which broadened lines result in severely overlapped cross-peaks.

Here we have used the *T. thermophilus* HB8 TTHA1718 gene product—a putative heavy-metal binding protein consisting of 66 amino acids that was overexpressed in *E. coli* to a concentration of 3−4 mM—as a model system. First we examined the stability of the live *E. coli* samples under measurement conditions at 37 °C. The experimental scheme of our in-cell NMR experiments is presented in Fig. 1a. The virtual identity of ^{1}H–^{15}N heteronuclear single quantum coherence (HSQC) spectra recorded immediately after sample preparation (Fig. 1b) and after 6 h in an NMR tube at 37 °C (Fig. 1c) show that TTHA1718 in-cell NMR samples are stable for at least 6 h. It is crucial for in-cell NMR to ensure that the proteins providing the NMR spectra are indeed inside the living cells, and that the contribution from extracellular proteins is negligible[15]. Most ^{1}H–^{15}N HSQC cross-peaks disappeared after removal of the bacteria by gentle centrifugation after the measurement shown in Fig. 1c (Fig. 1d), whereas the lysate spectrum of the collected cells shows much sharper cross-peaks (Supplementary Fig. 1b). These results were corroborated by SDS–PAGE (Supplementary Fig. 1c), demonstrating that the contribution of extracellular protein to the observed signals is negligible. The viability of the bacteria in the in-cell samples after 6 h of NMR measurements was confirmed to be 85 ± 11% by plating colony tests. TTHA1718 was indicated to be in the cytoplasm by cell fractionation experiments (Supplementary Fig. 1d, e).

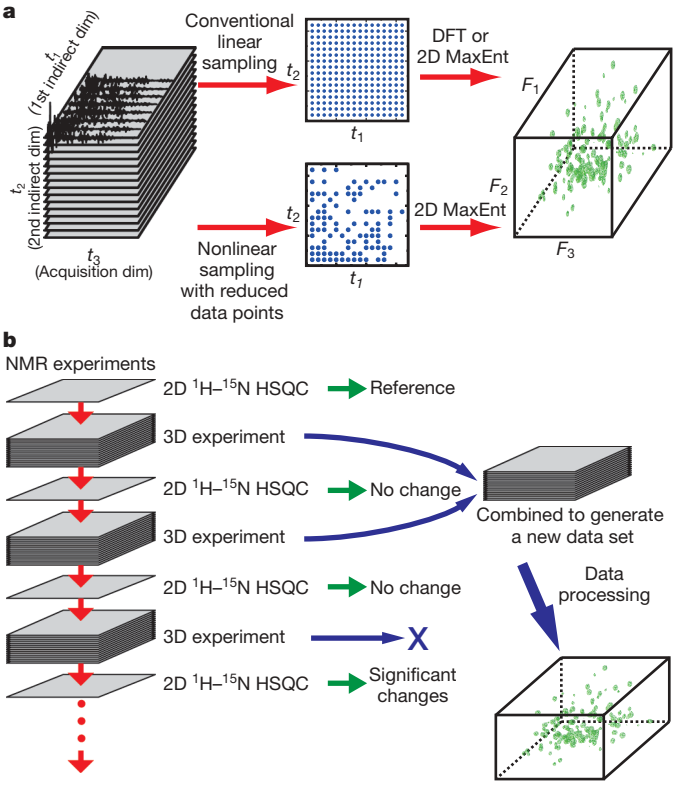

Figure 2 | Rapid acquisition of 3D NMR spectra of TTHA1718 in living *E. coli* cells. **a**, Rapid acquisition of 3D NMR spectra using a nonlinear sampling scheme. DFT, discrete Fourier transform; dim, dimension; MaxEnt, maximum-entropy processing. **b**, Repeated observation of 3D NMR spectra with intermittent monitoring of the sample condition by short 2D ^{1}H–^{15}N HSQC experiments.

Next we assigned backbone resonances of TTHA1718 in *E. coli* cells at 37 °C using six 3D triple-resonance NMR spectra. The short lifetimes of the in-cell NMR samples necessitated a large reduction in measurement times from the 1–2 days conventionally used for each 3D experiment. We therefore prepared a fresh sample for each experiment and applied a nonlinear sampling scheme for the indirectly acquired dimensions[16–18], which has been shown in combination with maximum-entropy processing to provide considerable time savings (Fig. 2a). With this technique the duration of each 3D experiment was reduced to 2–3 h. To ensure that only data of intact samples were acquired, each 3D experiment was repeated several times interleaved with monitoring of the sample condition by a short 2D ^{1}H–^{15}N HSQC experiment. These 3D data were combined to generate a new data set with improved signal-to-noise ratio until the 2D spectra showed marked changes (Fig. 2b). Typically, two 3D data sets were combined. All the expected backbone resonances were assigned except for Cys 11, Asn 12 and His 13 in the predicted metal-binding loop (Supplementary Fig. 2a, b). For comparison, the backbone resonances were also assigned for the purified TTHA1718 protein *in vitro*, in which Cys 11 was also assigned. A comparison of the chemical shifts under both conditions

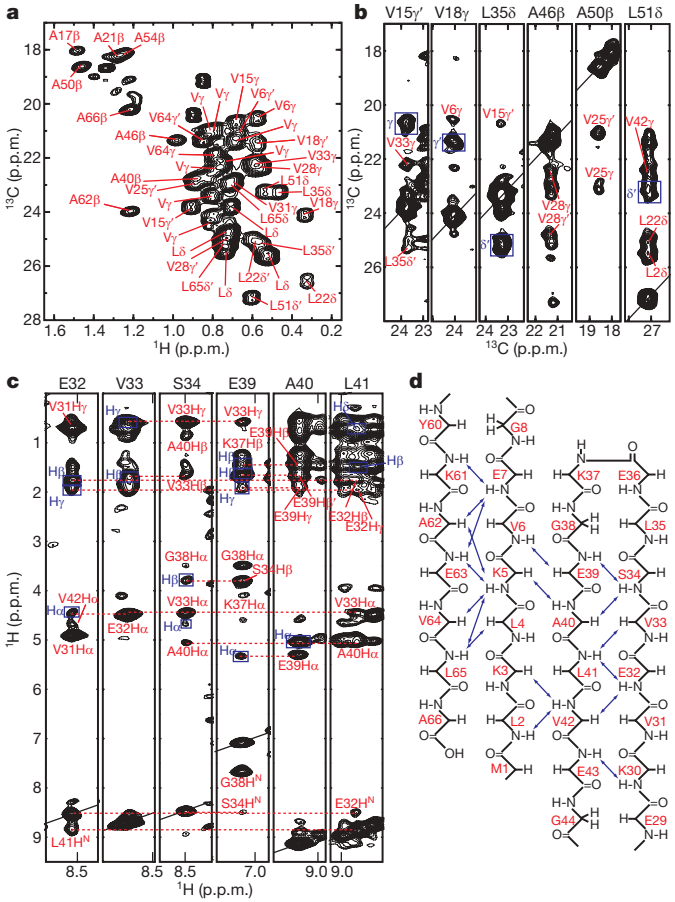

Figure 3 | Collection of NOE-derived distance restraints in TTHA1718 in living *E. coli* cells. **a**, Methyl region of the ^{1}H–^{13}C heteronuclear multiple-quantum coherence (HMQC) spectrum of the selectively methyl-protonated sample. Assignments of the methyl groups of Ala, Leu and Val residues are indicated, if available. **b**, ^{13}C–^{13}C cross-sections corresponding to the ^{1}H frequencies of representative methyl groups extracted from the 3D $^{13}C/^{13}C$-separated HMQC-NOE-HMQC spectrum. The cross-peaks due to interresidual NOEs are assigned in red. Intraresidual NOEs are indicated by blue boxes and annotated. **c**, ^{1}H–^{1}H cross-sections corresponding to the ^{15}N frequencies of selected backbone amide groups extracted from the 3D ^{15}N-separated NOESY-HSQC spectrum. The inter- and intraresidue NOEs are indicated as in **b**. **d**, Topology diagram of the β-sheet structure in TTHA1718. Interstrand backbone NOEs are depicted as double-headed arrows.

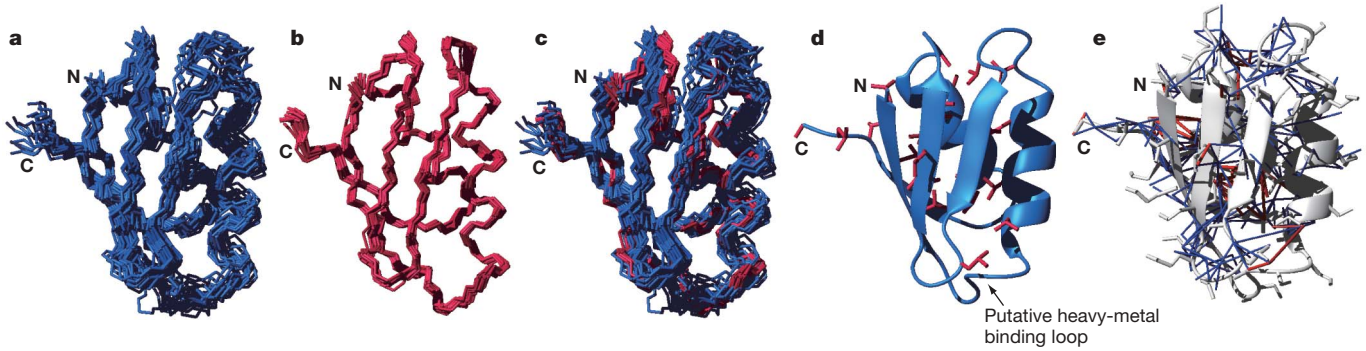

Figure 4 | NMR solution structure of TTHA1718 in living *E. coli* cells. **a**, A superposition of the 20 final structures of TTHA1718 in living *E. coli* cells, showing the backbone (N, Cα, C′) atoms. **b**, A superposition of the 20 final structures of purified TTHA1718 *in vitro*. **c**, A comparison of TTHA1718 structures in living *E. coli* cells and *in vitro*. The best fit superposition of backbone (N, Cα, C′) atoms of the two conformational ensembles are shown with the same colour code in **a** and **b**. **d**, Secondary structure of TTHA1718 in living *E. coli* cells. The side chains of Ala, Leu and Val residues, the methyl groups of which were labelled with $^{1}H/^{13}C$, are shown in red. **e**, Distance restraints derived from methyl-group-correlated and other NOEs are represented in the ribbon model with red and blue lines, respectively.

shows the largest differences for residues adjacent to the region with missing $^{1}H–^{15}N$ correlation peaks in the in-cell spectrum (Supplementary Fig. 2d). The chemical shifts of 86% of Hα, 71% of Hβ, and 34% of the other aliphatic $^{1}H/^{13}C$ side-chain resonances of TTHA1718 in *E. coli* cells were also determined (Supplementary Fig. 2c).

Assignments to side-chain methyl groups have a large effect on the structure calculation[19]. Methyl protonation at methionine residues has previously been used as probes in in-cell NMR[20]. We performed NMR measurements of TTHA1718 in *E. coli* cells selectively $^{1}H/^{13}C$-labelled at methyl groups[21] of Ala, Leu and Val (Supplementary Fig. 3), and assigned 31 out of 40 (78%) of their side-chain methyl ^{1}H and ^{13}C resonances (Fig. 3a). Overall, 148 NOEs involving methyl groups could be assigned (Fig. 3b) in the 3D nuclear Overhauser enhancement spectroscopy (NOESY) spectra and used in the structure calculation, including 69 out of all 89 long-range NOEs. Three-dimensional ^{15}N-separated and 3D ^{13}C-separated NOESY spectra measured on uniformly labelled *E. coli* samples yielded further NOE-derived distance restraints (Fig. 3c, d). In contrast to ^{15}N-labelling, uniform ^{13}C-labelling gave rise to a considerable number of 'background' cross-peaks (Supplementary Fig. 4)[20]. NOE cross-peaks in the 3D ^{13}C-separated NOESY were therefore carefully analysed and only selected if they were highly likely to correlate TTHA1718 resonances.

The 3D structure of TTHA1718 in *E. coli* cells was calculated with the program CYANA[22] on the basis of NOE distance restraints, backbone torsion angle restraints, and restraints for hydrogen bonds (Supplementary Table 1). The resulting structure is well-converged with a backbone root mean squared deviation (r.m.s.d.) of 0.96 Å to the mean coordinates (Fig. 4a), and is similar to the structure that was determined independently *in vitro* from a purified sample (Fig. 4b and Supplementary Table 1). The backbone r.m.s.d. between the in-cell and *in vitro* structures is 1.16 Å (Fig. 4c). Slight structural differences were found in the more dynamic loop regions, which may be due to the effects of viscosity and molecular crowding in the cytosol. In the putative heavy-metal binding loop, where the structural differences are corroborated by chemical shift differences, interactions with metal ions in the *E. coli* cytosol may affect the conformation of the region. Indeed, the C11S/C14S and C11A/C14A mutants, which lack metal-binding activity, showed almost identical $^{1}H–^{15}N$ HSQC spectra in *E. coli* cells even when the cells were loaded with excess metal ions such that for the wild-type protein, the Thr 10, Cys 14 and Val 15 resonances disappear and Met 9 experiences further chemical shift changes (Supplementary Figs 5–7).

Our results demonstrate that high resolution 3D structures of proteins can be determined in the cytoplasm of bacterial cells. Rapid data collection using nonlinear sampling and selective protonation at methyl groups (Fig. 4d, e) to enable the identification of unambiguous long-range NOE interactions was crucial for the success of this approach (Supplementary Fig. 8 and Supplementary Table 1).

The TTHA1718 protein was highly expressed in *E. coli* (3−4 mM in the NMR samples). Structural studies of less abundant proteins by in-cell NMR may however also be feasible, because 74% and 61% of the NOE cross-peaks used for the structure calculation could be still observed from in-cell NMR samples in which TTHA1718 expression level was controlled to approximately 1.2–1.5 and 0.6–0.8 mM ranges, respectively (Supplementary Fig. 9).

We have also tested the applicability of in-cell NMR to the structure determination of larger proteins by expressing rat calmodulin (17 kDa) in *E. coli* (concentration ~1.0–1.5 mM). Sequential backbone resonance connectivities and sequential $H^N–H^N$ NOEs could be identified in the 3D spectra (Supplementary Fig. 10), suggesting that structural analysis of this size of protein in *E. coli* cells will be feasible. In-cell NMR structure determination in eukaryotic cells should also be possible, for example in *Xenopus laevis* oocytes, in which proteins can be introduced at up to ~0.7 mM intracellular concentration[10], and here techniques developed for larger systems *in vitro*, for example, stereo-array isotope labelling[23] may be applied.

It has been proposed that the viscosity of the *E. coli* cytoplasm should be too high, and the tumbling of globular proteins therefore too slow to permit their NMR signals to be observed[24]. Apparently, this is not the case for TTHA1718 and calmodulin (see Supplementary Fig. 11 and Supplementary Table 2 for ^{15}N relaxation parameters of TTHA1718). However, it is likely that for some proteins, their interactions with other cellular components will markedly affect their rotational correlation times to the point that they will be effectively invisible to solution state NMR techniques. In-cell structure determination applied to proteins that adopt more than one extensively populated conformation, for example due to binding to multiple ligands, will pose additional problems.

Our results open new avenues for the investigation of protein conformations at atomic resolution and how they change in response to biological events in living environments. In particular, this approach provides the tools that will permit the effects of molecular crowding in the cytosol, the conformations of proteins that are intrinsically disordered *in vitro*, and the 3D structures of proteins that are otherwise unstable and difficult to purify to be investigated in living cells.

METHODS SUMMARY

E. coli cells harbouring the plasmid encoding the *T. thermophilus* HB8 TTHA1718 gene were first grown in unlabelled LB medium. Protein production was induced after transfer of the bacteria into stable isotope-labelled medium (100 ml). The collected cells were placed as ~60% slurry into NMR tubes.

Sample stability was monitored repeatedly by 2D ^{1}H–^{15}N HSQC spectra followed by plating colony tests. Samples for *in vitro* NMR experiments were purified by butyl-TOYOPEARL column chromatography. Sample preparation of TTHA1718 mutants and rat calmodulin was performed with essentially identical protocols. All NMR spectra were obtained at 37 °C using a Bruker Avance 600 spectrometer with Cryoprobe, processed using the AZARA 2.7 software (W. Boucher, www.bio.cam.ac.uk/azara), and analysed using the ANSIG 3.3 software[25]. NMR resonances of TTHA1718 in *E. coli* cells were assigned by analysing nine 3D triple-resonance NMR spectra. Intraresidue and sequential NOEs involving methyl protons were also used. Four 3D NOESY spectra were analysed for the collection of NOE-derived distance restraints. For all 3D NMR experiments, a nonlinear sampling scheme[16–18] was used for the indirectly observed dimensions to reduce experimental time. The 2D maximum-entropy method[26] was used for processing nonlinearly sampled dimensions. NMR analyses of purified TTHA1718 were performed using conventional approaches. TTHA1718 structures were calculated with the program CYANA using automated NOE assignment[27] and torsion angle dynamics[22]. Backbone torsion angle restraints from the program TALOS[28] were added to the input. Restraints were included for hydrogen bonds in regular secondary structure elements that were strongly supported by NOEs. The 20 final structures were embedded in a water shell and energy-minimised against the AMBER force field[29] with the program OPALp[30] in the presence of the experimental restraints.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 September 2008; accepted 22 January 2009.

1. Serber, Z. *et al.* High-resolution macromolecular NMR spectroscopy inside living cells. *J. Am. Chem. Soc.* **123,** 2446–2447 (2001).
2. Serber, Z., Corsini, L., Durst, F. & Dötsch, V. In-cell NMR spectroscopy. *Methods Enzymol.* **394,** 17–41 (2005).
3. Serber, Z. *et al.* Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nature Protocols* **1,** 2701–2709 (2006).
4. Reckel, S., Hänsel, R., Löhr, F. & Dötsch, V. In-cell NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **51,** 91–101 (2007).
5. Selenko, P. & Wagner, G. Looking into live cells with in-cell NMR spectroscopy. *J. Struct. Biol.* **158,** 244–253 (2007).
6. Ellis, R. J. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* **26,** 597–604 (2001).
7. Burz, D. S., Dutta, K., Cowburn, D. & Shekhtman, A. Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nature Methods* **3,** 91–93 (2006).
8. McNulty, B. C., Young, G. B. & Pielak, G. J. Macromolecular crowding in the *Escherichia coli* periplasm maintains α-synuclein disorder. *J. Mol. Biol.* **355,** 893–897 (2006).
9. Dedmon, M. M., Patel, C. N., Young, G. B. & Pielak, G. J. FlgM gains structure in living cells. *Proc. Natl Acad. Sci. USA* **99,** 12681–12684 (2002).
10. Selenko, P., Serber, Z., Gade, B., Ruderman, J. & Wagner, G. Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis* egg extracts and intact oocytes. *Proc. Natl Acad. Sci. USA* **103,** 11904–11909 (2006).
11. Sakai, T. *et al.* In-cell NMR spectroscopy of proteins inside *Xenopus laevis* oocytes. *J. Biomol. NMR* **36,** 179–188 (2006).
12. Inomata, K. *et al.* High-resolution multi-dimensional NMR spectroscopy of proteins in living human cells. *Nature* doi:10.1038/nature07839 (this issue).
13. Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley, 1986).
14. Reardon, P. N. & Spicer, L. D. Multidimensional NMR spectroscopy for protein characterization and assignment inside cells. *J. Am. Chem. Soc.* **127,** 10848–10849 (2005).
15. Serber, Z., Ledwidge, R., Miller, S. M. & Dötsch, V. Evaluation of parameters critical to observing proteins inside living *Escherichia coli* by in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **123,** 8895–8901 (2001).
16. Barna, J. C. J., Laue, E. D., Mayger, M. R., Skilling, J. & Worrall, S. J. P. Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *J. Magn. Reson.* **73,** 69–77 (1987).
17. Schmieder, P., Stern, A. S., Wagner, G. & Hoch, J. C. Improved resolution in triple-resonance spectra by nonlinear sampling in the constant-time domain. *J. Biomol. NMR* **4,** 483–490 (1994).
18. Rovnyak, D. *et al.* Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *J. Magn. Reson.* **170,** 15–21 (2004).
19. Mueller, G. A. *et al.* Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein. *J. Mol. Biol.* **300,** 197–212 (2000).
20. Serber, Z. *et al.* Methyl groups as probes for proteins and complexes in in-cell NMR experiments. *J. Am. Chem. Soc.* **126,** 7119–7125 (2004).
21. Rosen, M. K. *et al.* Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol.* **263,** 627–636 (1996).
22. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273,** 283–298 (1997).
23. Kainosho, M. *et al.* Optimal isotope labelling for NMR protein structure determinations. *Nature* **440,** 52–57 (2006).
24. Li, C. *et al.* Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: implications for in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **130,** 6310–6311 (2008).
25. Kraulis, P. J., Domaille, P. J., Campbell-Burk, S. L., Van Aken, T. & Laue, E. D. Solution structure and dynamics of Ras p21-GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry* **33,** 3515–3531 (1994).
26. Laue, E. D., Mayger, M. R., Skilling, J. & Staunton, J. Reconstruction of phase sensitive 2D NMR spectra by maximum entropy. *J. Magn. Reson.* **68,** 14–29 (1986).
27. Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319,** 209–227 (2002).
28. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13,** 289–302 (1999).
29. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117,** 5179–5197 (1995).
30. Koradi, R., Billeter, M. & Güntert, P. Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun.* **124,** 139–147 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank S. Kuramitsu for providing the plasmid encoding TTHA1718, and D. Nietlispach for setting up 3D NMR experiments with nonlinear sampling schemes and ^{15}N relaxation experiments, T. Anzai for assistance with NMR measurements, and H. Koyama and A. Iwasaki for sample preparations. This work was supported in part by CREST, Japan Science and Technology Agency (JST), the Molecular Ensemble Program, RIKEN, Grants-in-Aid for Scientific Research of Priority Areas from the Japanese Ministry of Education, Sports, Culture, Science, and Technology on 'Molecular Soft Interactions Regulating Membrane Interface of Biological Systems' and 'Molecular Science for Supra Functional Systems – Development of Advanced Methods for Exploring Elementary Process', and by the Volkswagen Foundation.

Author Contributions B.O.S., M.S., P.G. and Y.I. designed the research and wrote the manuscript. D.S., A.S. and T.I. conducted the research including sample preparation, data acquisition, resonance assignment and structure calculation. M.M. and M.W. helped with NMR measurements. M.M. prepared TTHA1718 mutants. J.H. and T.H. measured NMR data on TTHA1718 mutants and ^{15}N-relaxation experiments. N.H. provided the expression vector for calmodulin. M.Y. measured NMR data on calmodulin in living *E. coli* cells. T.M. helped during the preparation and characterisation of TTHA1718.

Author Information Atomic coordinates of the structures of TTHA1718 in *E. coli* cells and *in vitro* have been deposited in the Protein Data Bank under accession codes 2ROG and 2ROE, respectively. Chemical shifts have been deposited in the BioMagResBank under accession numbers 11037 and 11035. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Y.I. (ito-yutaka@tmu.ac.jp).

METHODS

Sample preparation. The expression plasmid encoding the *T. thermophilus* HB8 TTHA1718 gene was obtained from the 'Whole-Cell Project of a Model Organism, *T. thermophilus* HB8' (http://www.thermus.org). In-cell NMR samples were prepared as follows. JM109 (DE3) *E. coli* cells harbouring the TTHA1718 expression plasmid were first grown in unlabelled LB medium. The production of uniformly $^{13}C/^{15}N$-labelled TTHA1718 was induced by the addition of isopropyl thio-β-D-thiogalactoside to a final concentration of 0.5 mM after transfer of the bacteria into M9 minimal medium (100 ml) containing 2 g l^{-1} [U-^{13}C]-glucose and 1 g l^{-1} $^{15}NH_4Cl$. For the production of TTHA1718 samples with selectively protonated side-chain methyl groups of Ala, Leu and Val residues in a uniform ^{2}H-background, protein expression was induced in 100% D_2O M9 medium containing 2 g l^{-1} unlabelled glucose, 1 g l^{-1} $^{15}NH_4Cl$, 100 mg l^{-1} [3-^{13}C]-alanine and 100 mg l^{-1} [U-^{13}C, 3-^{2}H]-α-ketoisovalerate. For Val/Leu selective methyl protonation, [3-^{13}C]-alanine was excluded from the medium, whereas unlabelled leucine (100 mg l^{-1}) was added for Ala/Val selective protonation. The cells were collected by gentle centrifugation and placed as ~60% slurry with M9 medium containing 10% D_2O into NMR tubes. The concentration of TTHA1718 in *E. coli* samples was estimated by comparing the density of the Coomassie-stained bands in SDS–PAGE gels with those of proteins with similar molecular size and known concentration. The stability of TTHA1718 *E. coli* samples was monitored repeatedly by 2D ^{1}H–^{15}N HSQC spectra followed by plating colony tests. The localization of TTHA1718 in *E. coli* cells was first predicted from its amino acid sequence by PSORTb v.2.0 (refs 31, 32) (http://www.psort.org/psortb/) and SignalP 3.0 (refs 33, 34) (http://www.cbs.dtu.dk/services/SignalP/). The localization of overexpressed TTHA1718 was then analysed by measuring 2D ^{1}H–^{15}N HSQC spectra of spheroplasts and periplasmic extract, which were fractionated from TTHA1718-expressing ^{15}N-labelled cells by lysozyme-EDTA treatment using the conditions described previously[35]. Spheroplast formation was monitored by light microscopy.

TTHA1718 proteins for *in vitro* NMR experiment was purified by butyl-TOYOPEARL column chromatography after cell lysis, by sonication and high temperature (70 °C) treatment for 10 min. The final TTHA1718 fractions were concentrated and dissolved in M9 medium containing 10% D_2O for NMR experiments.

Two cysteine residues (C11 and C14) were predicted to be responsible for metal binding activity on the basis of analysis of multiple sequence alignment with previously characterized homologous proteins. Two double mutations (C11S/C14S and C11A/C14A) were therefore introduced by site-directed mutagenesis to disrupt metal binding.

Rat calmodulin in-cell NMR samples were produced similarly from the expression plasmid[36].

NMR spectroscopy. NMR experiments were performed at 37 °C probe temperature in a triple-resonance cryoprobe fitted with a *z*-axis pulsed field gradient coil, using a Bruker Avance 600 MHz spectrometer. NMR spectra were processed using the AZARA 2.7 software (W. Boucher, http://www.bio.cam.ac.uk/azara/), and analysed using an OpenGL version of the ANSIG 3.3 software[25,37].

Backbone $^{1}H^{N}$, ^{15}N, $^{13}C\alpha$, $^{13}C'$, and side-chain $^{13}C\beta$ resonance assignments of TTHA1718 in living *E. coli* cells were performed by analysing six 3D triple-resonance NMR spectra: HNCA, HN(CO)CA, CBCA(CO)NH, CBCANH, HNCO and HN(CA)CO. Three-dimensional HBHA(CBCACO)NH, H(CCCO)NH and (H)CC(CO)NH experiments were performed for side-chain ^{1}H and ^{13}C resonance assignments. Longitudinal (T_1) and transverse (T_2) ^{15}N relaxation parameters of TTHA1718 in living *E. coli* cells were obtained by measuring 1D ^{15}N-edited ^{15}N T_1 or T_2 relaxation experiments on lysine-selectively ^{15}N-labelled samples. Intraresidue and sequential NOEs involving methyl protons were also used for the assignment of Ala/Leu/Val methyl groups. ^{1}H–^{13}C HMQC spectra of in-cell NMR samples with three different methyl-selective labelling patterns, Ala/Val, Leu/Val, and Ala/Leu/Val, were used for amino acid classification of methyl ^{1}H–^{13}C correlation cross-peaks. For the collection of NOE-derived distance restraints, 3D ^{15}N-separated and ^{13}C-separated NOESY-HSQC spectra were measured on uniformly labelled in-cell NMR samples. In addition, 3D ^{13}C-separated NOESY-HSQC and 3D $^{13}C/^{13}C$-separated HMQC-NOESY-HMQC spectra were measured on Ala/Leu/Val-methyl-selectively protonated samples. For all 3D NMR experiments, a non-linear sampling scheme[16–18] was used for the indirectly observed dimensions to reduce experimental time. In brief, approximately one-quarter of the points were selected in a pseudo-random fashion from the conventional regularly spaced grid of t_1, t_2 points. The 2D maximum-entropy method[26] was used for processing nonlinearly sampled dimensions.

Backbone and side-chain resonances of purified TTHA1718 were assigned by analysis of seven 3D triple-resonance experiments: CBCA(CO)NH, CBCANH, HNCO, HN(CA)CO, H(CCCO)NH, (H)CC(CO)NH and HCCH-TOCSY, and almost completely assigned. NOE-derived distance restraints were obtained by analysing 3D ^{15}N-separated and ^{13}C-separated NOESY-HSQC spectra.

Structure calculation. TTHA1718 structures were calculated with the program CYANA[38] version 3.0 using automated NOE assignment[27] and torsion angle dynamics for the structure calculation[22], which was started from 100 conformers with random torsion angle values. The standard CYANA simulated annealing schedule was applied with 10,000 torsion angle dynamics steps. Backbone torsion angle restraints obtained from chemical shifts with the program TALOS[28] were added to the input for CYANA. Distance restraints for hydrogen bonds were introduced for the particular positions in the β-sheet region where the existences of hydrogen bonds were strongly suggested by inter-strand NOEs. Some additional 'CYANA-estimated' hydrogen bond restraints were included during the calculation process. The 20 conformers with the lowest final CYANA target function values were embedded in a water shell of 8 Å thickness and energy-minimised against the AMBER force field[29] with the program OPALp[30] in the presence of the experimental restraints.

31. Gardy, J. L. *et al.* PSORT-B: Improving protein subcellular localization prediction for gram-negative bacteria. *Nucleic Acids Res.* **31,** 3613–3617 (2003).
32. Gardy, J. L. *et al.* PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **21,** 617–623 (2005).
33. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10,** 1–6 (1997).
34. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340,** 783–795 (2004).
35. Thorstenson, Y. R., Zhang, Y., Olson, P. S. & Mascarenhas, D. Leaderless polypeptides efficiently extracted from whole cells by osmotic shock. *J. Bacteriol.* **179,** 5333–5339 (1997).
36. Hayashi, N., Matsubara, M., Takasaki, A., Titani, K. & Taniguchi, H. An expression system of rat calmodulin using T7 phage promoter in Escherichia coli. *Protein Expr. Purif.* **12,** 25–28 (1998).
37. Kraulis, P. J. ANSIG: a program for the assignment of protein ^{1}H 2D NMR spectra by interactive computer graphics. *J. Magn. Reson.* **84,** 627–633 (1989).
38. Güntert, P. Automated NMR protein structure calculation. *Prog. Nucl. Magn. Reson. Spectrosc.* **43,** 105–125 (2003).

NEWS & VIEWS

STRUCTURAL BIOLOGY

Inside the living cell

David S. Burz and Alexander Shekhtman

Proteins work properly only if they have the correct three-dimensional atomic structure. It is now possible to look at the structures and dynamics of these biological macromolecules as they function inside cells.

The development of structural biology was historically based on the principle of divide and conquer — individual proteins were purified to homogeneity and their atomic structures were solved *in vitro* by using either X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. This approach was tremendously successful, and led to the creation of a protein-structure databank that currently contains more than 50,000 structures.

But relating *in vitro* protein structures to biological processes that occur inside the cell is not a trivial task. A traditional approach to solving this problem entails mutating a protein's structure at certain sites based on its *in vitro* structure and observing the effects of these changes on the cell. This low-resolution validation of high-resolution structures may still lead to situations where the *in vitro* structure does not fully represent the physiologically active protein structure under the conditions present in a cell. The work presented by Sakakibara *et al.*[1] and Inomata *et al.*[2] (pages 102 and 106 of this issue) reveals new ways to solve the structures of proteins as they exist inside living cells, ushering in a fresh era of structural biology.

To determine how protein structures are influenced by their intracellular environment, in-cell NMR spectroscopy was developed[3]. NMR spectroscopy allows one to directly observe NMR-active isotopes of atomic nuclei within any NMR-inactive environment, and can thus be used to analyse isotopically labelled biomolecules inside unlabelled cells. To date, two approaches have been used to deliver labelled proteins into unlabelled bacterial and animal cells. In the first case, target proteins are produced inside the bacteria by growing them on isotopically labelled media; in the second, labelled proteins are microinjected into large cells such as *Xenopus* oocytes (frog eggs)[4]. In these instances, in-cell NMR spectra suggested that protein structures inside cells are very similar to those solved *in vitro*. The devil, though, is in the detail.

Changes in protein structure that are caused by specific interactions with well-defined binding partners can be identified by solving *in vitro* structures of the protein complexes. A more difficult problem is to address how the numerous, nonspecific, low-affinity interactions, which are omnipresent in cells, affect a protein's structure. Living cells are extremely crowded environments owing to the high (20–30% by volume) concentration of macromolecules they contain, and this results in small but potentially important changes in protein structure. The papers published in this issue[1,2] make significant steps towards understanding these changes.

A major hurdle to determining in-cell NMR structure is the limited lifetime of the cells inside the NMR sample tube. Standard NMR experiments usually require 1–2 days of data collection, which is an unacceptably long time for live cells. Sakakibara *et al.*[1] shortened this time to 2–3 hours by applying a well-known but seldom-used modification of NMR experiments, and thereby determined the three-dimensional structure of a putative heavy-metal-binding protein, TTHA1718, expressed inside bacterial cells (Fig. 1a). Their procedure, as described in detail in the paper, may well become a new standard for in-cell NMR.

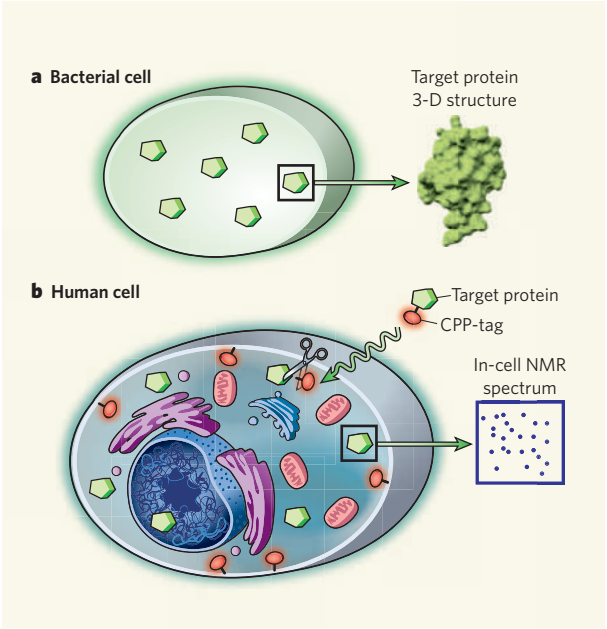

Figure 1 | Analysis of proteins using in-cell NMR spectroscopy. **a**, Sakakibara *et al.*[1] show that atomic structures of proteins expressed inside bacterial cells can be solved by in-cell NMR spectroscopy. **b**, Inomata *et al.*[2] fused a target protein with a cell-penetrating peptide (CPP-tag) to transfer stable-isotope-labelled proteins into various human cells and collect in-cell NMR spectra at atomic resolution. This approach has not yet been used to determine protein structures but has been applied by the authors to study protein dynamics.

Comparing the *in vitro* and in-cell structures of TTHA1718 revealed that, despite marked similarities, there are structural differences, mostly concentrated in the heavy-metal-binding site and in loop regions that undergo dynamic changes as the protein functions. In contrast to the changes in the binding site, which can be explained by metal ions present in the bacterial cytosol, the structural changes seen in the dynamic loop are probably due to molecular crowding and the viscosity of the cytosol that are characteristic of the cell interior. It will be interesting to discover whether this phenomenon is seen when further in-cell NMR structures become available.

Extending in-cell NMR to study proteins inside human cells presents a further challenge. In general, protein production inside these cells does not reach high enough levels for atomic-resolution in-cell NMR spectra to be collected. The microinjection technique is laborious, and up to now has been limited to large cells such as *Xenopus* oocytes. Inomata *et al.*[2] describe an innovative method that avoids using microinjection and makes in-cell NMR in human cells possible. By fusing the labelled target protein with a cell-penetrating peptide (Fig. 1b), the target protein can be delivered into the cells, where the peptide tag is then snipped off, allowing the free, labelled target protein to disperse uniformly. Importantly,

the authors found that the tag binds to large intracellular structure(s), making it invisible to NMR and thus simplifying the in-cell NMR spectrum of the target protein.

This method opens the door to determining protein structures inside human cells in the near future, but has already been used by Inomata *et al.* to study in-cell protein dynamics. For example, macromolecular crowding inside the cell should stabilize protein structures. Inomata *et al.* show that, for at least one protein, ubiquitin, introduced into a human cell at physiological concentrations, the opposite is true. Ubiquitin becomes more dynamic in its reactivity and less structured, presumably due to nonspecific, low-affinity interactions with its binding partners. This unexpected result highlights the importance of studying proteins inside living cells.

In-cell NMR is limited by the concentration and structural stability of the protein that can be attained inside cells. In addition, some proteins could be difficult to deliver into the cytosol by fusing them with cell-penetrating peptides. Very large proteins, or proteins that bind to large cellular structures, become invisible and cannot be effectively studied by in-cell NMR. The lifetime of the cells is also a limiting factor, because cell breakdown results in protein leaking away into solution, the condition of *in vitro* NMR. Despite these pitfalls, we are left with an enormous number of proteins that can be studied.

Exploring protein structures and dynamics at the atomic level inside living cells will provide results that cannot be obtained using standard *in vitro* techniques. The comparative simplicity of the in-cell method allows for a myriad of applications. The regulation of metabolic or signal-transduction pathways, mediated by biomolecular interactions, can now be studied in detail. Drug screening using in-cell NMR can function as an *in vivo* assay at atomic resolution, providing information about drug delivery inside the cell, where the drug binds, and whether there is a notable difference between how it binds *in vivo* and *in vitro*. More exotic applications include the study of intrinsically unstructured and amyloid-forming proteins in neurons — such proteins having been implicated in neurodegeneration — or of labelled protein probes in diseased tissue. The ability to observe the structures of proteins in their native environment releases the constraints that have previously limited study of protein structure and dynamics to the test tube. Now that structural biology has moved into the cell, it is likely to stay there. ■

David S. Burz and Alexander Shekhtman are in the Department of Chemistry, State University of New York at Albany, 1400 Washington Avenue, Albany, New York 12222, USA.
e-mail: ashekhta@albany.edu

1. Sakakibara, D. *Nature* **458,** 102–105 (2009).
2. Inomata, K. *et al. Nature* **458,** 106–109 (2009).
3. Serber, Z. *et al. J. Am. Chem. Soc.* **123,** 2446–2447 (2001).
4. Selenko, P. *et al. Proc. Natl Acad. Sci. USA* **103,** 11904–11909 (2006).

CONDENSED-MATTER PHYSICS

Carbon conductor corrupted

Michael S. Fuhrer and Shaffique Adam

Atomically thin sheets of graphite are metal-like conductors — until they react with hydrogen, when they become insulators. This curious effect could be an excellent model for studying metal–insulator transitions.

In most solids, electrons behave much like particles of matter: they have a mass, and they speed up and slow down in response to forces. But in graphene — the single-atom-thick sheet of carbon that constitutes the basic building block of graphite — electrons move as if they have no mass[1,2], and so behave more like photons. In other words, although electrons in graphene can change their momentum and energy, they cannot speed up or slow down. One would therefore intuitively think that electron flow (electrical current) in graphene could never be completely blocked. But reporting in *Science*, Elias *et al.*[3] show that, when graphene reacts with a small amount of hydrogen, its electrons become stuck and the carbon sheet becomes an insulator.

The band structure of a material describes which energy states can be occupied by the material's constituent electrons. Just like other fermions, electrons in a solid fill up the lowest-energy bands first, before filling up bands at higher energies, like water filling a bath. If electrons at the top surface — the Fermi level — of the resulting electron sea can slosh about into the unoccupied part of the band structure, then the solid is said to be a metal, in which electrical currents are created by electrons moving from one momentum state to another (Fig. 1a). But if the electrons are not free to move because of a gap in the band structure, then the solid is a 'band insulator' (Fig. 1b).

There is, however, another way to make an insulator. Electrons are quantum-mechanical objects that behave as waves. Constructive interference of electron waves near any imperfections — disorder — in a solid creates standing waves (Fig. 1c) that are localized in real space, effectively 'freezing' the electron sea[4]. A long-standing conjecture in physics states that electrons in any two-dimensional system will become localized by any disorder, no matter how small. Because all two-dimensional electronic systems contain some disorder, all such systems should therefore be insulators[5]. In practice, however, very low temperatures and/or large samples are needed to reveal this insulating behaviour, and even then the effect is not always seen.

So how does graphene fit into all this? The band structure of graphene can be thought of as a cone balancing on its tip, atop the point of another cone (Fig. 1d). Because there is no bandgap, graphene is a metallic conductor. But graphene differs from other metals when its Fermi level lies at the Dirac point — the point where the cones touch. There, the top surface of the electron sea becomes vanishingly small. One consequence is that, in contrast to disorder-free metals (which would have infinite conductivity), perfect sheets of graphene are expected to have a conductivity of $4e^2/\pi h$, where e is the electronic charge, and h is Planck's constant. Furthermore, graphene is thought to be the only exception to the localization conjecture: if it contains only 'smooth' disorder in which there are no sharp changes between neighbouring carbon atoms, then graphene remains metallic, and even quantum interference cannot localize its electrons[6].

In reality, graphene is found to have a sample-dependent, finite minimum conductivity that is always greater than $4e^2/\pi h$ (ref. 1), thus creating a mystery — how could the disorder present in real graphene increase the conductivity above the theoretical value for perfect graphene? Disorder normally increases the scattering of electrons, which decreases conductivity.

The answer lies in the nature of the disorder. Most graphene samples are dirtied by charged impurities that lie near the graphene sheet, either on the surface or in the nearby substrate[7]. These charges have two effects: they create a smooth disorder that scatters electrons (which reduces conductivity), but they also either attract or repel electrons, creating local electron-rich or electron-poor 'puddles' whose Fermi levels lie above or below the Dirac point. The electrons in these puddles increase the conductivity of the sample, counteracting the decrease due to scattering[8].

By adding hydrogen to graphene, Elias *et al.*[3] were able to study a fundamentally different sort of disorder. By reacting all the carbon atoms in a graphene sheet with hydrogen, so that each carbon becomes bonded to a single hydrogen atom, the authors made a new kind of two-dimensional crystal. This material, known as graphane, is expected to be a conventional band insulator[9].

But at lower doses, hydrogen probably bonds randomly to only a few carbon atoms. Unlike charged impurities, which cause a smooth disorder, hydrogen creates sharply varying disorder, because a carbon atom bonded to hydrogen is very different from its neighbours. The authors[3] observed that their partially hydrogenated graphene had greatly reduced minimum conductivity, which varied with temperature and tended towards zero at low temperatures — the signature of an insulator. This is in contrast to graphene containing