# NMR protein structure determination in living *E. coli* cells using nonlinear sampling

Teppei Ikeya[1,2], Atsuko Sasaki[1,3], Daisuke Sakakibara[1,3], Yoshiki Shigemitsu[1,3], Junpei Hamatsu[1,3], Tomomi Hanashima[1], Masaki Mishima[1,3], Masatoshi Yoshimasu[4], Nobuhiro Hayashi[5], Tsutomu Mikawa[6,7], Daniel Nietlispach[8], Markus Wälchli[9], Brian O Smith[10], Masahiro Shirakawa[3,11], Peter Güntert[1,2,12] & Yutaka Ito[1,3,6,7]

[1]Department of Chemistry, Tokyo Metropolitan University, Hachioji, Tokyo, Japan. [2]Center for Biomolecular Magnetic Resonance, Institute of Biophysical Chemistry, Goethe University, Frankfurt am Main, Germany. [3]CREST, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan. [4]Cellular and Molecular Biology Laboratory, RIKEN, Wako-shi, Saitama, Japan. [5]Department of Life Science, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama, Kanagawa, Japan. [6]Research Group for Bio-Supramolecular Structure-Function, RIKEN, Yokohama, Kanagawa, Japan. [7]Biometal Science Laboratory, RIKEN, Sayo-gun, Hyogo, Japan. [8]Department of Biochemistry, University of Cambridge, Cambridge, UK. [9]Bruker BioSpin, Yokohama, Kanagawa, Japan. [10]Division of Molecular and Cellular Biology, Faculty of Biomedical and Life Sciences, University of Glasgow, Glasgow, UK. [11]Department of Molecular Engineering, Graduate School of Engineering, Kyoto University, Nishikyo-Ku, Kyoto, Japan. [12]Frankfurt Institute for Advanced Studies, Frankfurt am Main, Germany. Correspondence should be addressed to Y.I. (ito-yutaka@tmu.ac.jp).

Published online 13 May 2010; doi:10.1038/nprot.2010.69

**The cell is a crowded environment in which proteins interact specifically with other proteins, nucleic acids, cofactors and ligands. Atomic resolution structural explanation of proteins functioning in this environment is a main goal of biochemical research. Recent improvements to nuclear magnetic resonance (NMR) hardware and methodology allow the measurement of high-resolution heteronuclear multidimensional NMR spectra of macromolecules in living cells (in-cell NMR). In this study, we describe a protocol for the stable isotope ($^{13}C$, $^{15}N$ and $^{2}H$) labeling and structure determination of proteins overexpressed in *Escherichia coli* cells exclusively on the basis of information obtained in living cells. The protocol combines the preparation of the protein in *E. coli* cells, the rapid measurement of the three-dimensional (3D) NMR spectra by nonlinear sampling of the indirectly acquired dimensions, structure calculation and structure refinement. Under favorable circumstances, this in-cell NMR approach can provide high-resolution 3D structures of proteins in living environments. The protocol has been used to solve the first 3D structure of a protein in living cells for the putative heavy metal-binding protein TTHA1718 from *Thermus thermophilus* HB8 overexpressed in *E. coli* cells. As no protein purification is necessary, a sample for in-cell NMR measurements can be obtained within 2–3 d. With the nonlinear sampling scheme, the duration of each 3D experiment can be reduced to 2–3 h. Once chemical shift assignments and NOESY peak lists have been prepared, structure calculation with the program CYANA and energy refinement can be completed in less than 1 h on a powerful computer system.**



## INTRODUCTION

There are many widely used methods for determining the 3D structure of purified proteins in single crystals or in solution, which have resulted in very valuable contributions to the understanding of many biological processes. It is, however, very difficult to replicate the cellular environment *in vitro* and it is an interesting open question whether the extreme crowding by macromolecules in cells[1] influences the behavior of proteins. Therefore, *in vivo* information regarding the 3D structures, dynamics and interactions of proteins is required to better understand the structural basis of their functions inside cells. The noninvasive character of NMR spectroscopy and its ability to provide data at atomic resolution make NMR ideally suitable for the task[2]. We have recently developed a method for investigating protein structures in *E. coli* cells by combining a range of existing NMR techniques and computer programs. The approach has been used to characterize the structure of the *Thermus thermophilus* HB8 protein TTHA1718 in living *E. coli* cells[3]. In this article, we describe the protocol of our method that includes sample preparation, NMR measurements, resonance assignment, NOESY assignment, structure calculation and structure refinement.

### In-cell NMR

The advantages of NMR spectroscopy have been combined to obtain information on the conformation and dynamics of biological macromolecules inside living cells[4–10]. In-cell NMR experiments require labeling of target proteins with NMR-active stable isotopes such as $^{13}C$ and $^{15}N$ inside host cells. There were two approaches for achieving this condition. The first approach, which has been used for bacterial in-cell NMR studies[11–13], is to prepare target proteins using intrinsic protein expression systems in host cells. The second approach, which has been applied to *Xenopus* oocytes (frog eggs)[14–16], is to use purified proteins and incorporate them into cells by microinjection. It is even more difficult to study proteins inside cultured human cells, because the concentrations of individual proteins are too low for in-cell NMR spectra to be collected, and the artificial delivery of labeled proteins by microinjection is limited to extraordinarily large cells such as oocytes. An alternative method enables in-cell NMR in human cells by tagging a labeled target protein with a cell-penetrating peptide[17]. The tag is cleaved off in the cell, and the target protein is released. The cleaved tag becomes invisible to NMR because it binds to large intracellular structures. Recently, another method for mammalian in-cell NMR was reported, in which labeled proteins were introduced into cells through resealable pores formed by streptolysin O toxin[18].

### Protein structure determination inside cells

Until recently, the low sensitivity and short lifetime of samples have prevented the acquisition of sufficient structural information to determine protein structures by in-cell NMR. Determining the 3D structures of proteins inside cultured human cells is not feasible yet, as the intracellular concentration of the introduced target

proteins achieved so far is at the most ~30 μM[17]. It may be possible to determine the structures of proteins in *X. laevis* oocytes or eggs, into which proteins were injected at up to ~0.7 mM intracellular concentration[14], thus fulfilling the requirement on protein concentration for NOESY-type experiments. Bacterial in-cell NMR is the easiest system for investigating the effect of molecular crowding on protein structures. We therefore focus on an *E. coli*-based in-cell NMR method for 3D protein structure determination. *De novo* NMR protein structure determination in living *E. coli* cells requires methods for resonance assignment that do not rely on information obtained *in vitro*, and for obtaining distance information from nuclear Overhauser effects (NOEs), even in cases in which the broadened lines observed from in-cell samples can result in severely overlapped cross-peaks. These problems were overcome with the first 3D protein structure calculated exclusively on the basis of information obtained in living *E. coli* cells[3].

One important problem is that the length of time required to collect sufficient data to determine an in-cell protein structure by NMR is longer than the time that cells can remain viable in an NMR tube. This can be overcome, however, by preparing a fresh sample for each experiment and shortening their durations to 2–3 h by applying a nonlinear sampling scheme for the indirectly acquired dimensions[19–21]. This makes it possible for the requisite 3D NMR experiments to be performed. As a result, the expected NMR resonances for the backbone and most of side-chain NMR resonances can be observed and assigned, and a sufficient number of NOE distance restraints can be collected using a sample with selectively $^{1}H/^{13}C$-labeled methyl groups of alanine, leucine and valine, as well as 3D $^{15}N$-separated NOESY and 3D $^{13}C$-separated NOESY spectra measured on uniformly labeled *E. coli* samples. Methyl-selective protonation has a large impact on the structural analysis of larger proteins[22], and has also been used to provide site-specific probes in in-cell NMR[23]. On the basis of these data, the 3D structure of the protein can be calculated with the program CYANA[24].

The viscosity inside cells[25] increases the rotational correlation time and apparent molecular mass of proteins. We expect that in-cell structure determination by our approach will be feasible for proteins up to ~20 kDa, as we could identify the sequential backbone resonance connectivities and sequential $H^N$-$H^N$ NOEs for rat calmodulin (17 kDa) in *E. coli* cells[3].

### Experimental design

To produce a sample of *E. coli* cells in which only the target protein is labeled with $^{13}C$ and $^{15}N$, we generally first grow cells harboring the expression plasmid in unlabeled LB medium, and then transfer into M9 minimal medium containing stable isotopes (see **Table 1**) in which protein expression is induced (e.g., by the addition of isopropyl β-D-1-thiogalactopyranoside). The timing of induction, the optimal temperature and the incubation times for the protein expression should be determined at small scale before large-scale experiments. We usually try at least two *E. coli* strains, JM109 (a K-12 strain) and BL21 (a B strain), when starting in-cell NMR experiments with a new protein. In many cases, we obtain similar results from both the strains. However, we also observed the case of rat calmodulin, in which, for reasons unknown, good $^{1}H$-$^{15}N$ heteronuclear single quantum coherence (HSQC) spectra were observed when using JM109, whereas broadened cross-peaks were observed when using BL21 (M.Y., N.H., T.M. and Y.I., unpublished results). For the collection of conformational restraints by in-cell NMR, more samples are generated with selectively protonated side-chain methyl groups of Ala, Leu and Val residues in a uniform $^{2}H$ background. The M9 media for Val/Leu, Ala/Val and Ala/Leu/Val selective methyl protonation are summarized in **Table 1**. Cells are harvested by gentle centrifugation and placed

**TABLE 1** | M9 minimal medium for $^{13}C/^{15}N$ uniform labeling (UL) and selective protonation at side-chain methyl groups.

| | $^{13}C/^{15}N$ UL | Ala/Leu/Val | Val/Leu | Ala/Val |
|---|---|---|---|---|
| $^{15}NH_4Cl$ | 0.1 g | 0.1 g | 0.1 g | 0.1 g |
| [U-$^{13}C$] glucose | 0.2 g | — | — | — |
| [3-$^{13}C$] alanine | — | 0.01 g | — | 0.01 g |
| [U-$^{13}C$, 3-$^{2}H$] α-ketoisovalerate | — | 0.01 g | 0.01 g | 0.01 g |
| Unlabeled glucose | — | 0.2 g | 0.2 g | 0.2 g |
| Unlabeled leucine | — | — | — | 0.01 g |
| $NH_4Cl$ | — | 0.1 g | 0.1 g | 0.1 g |
| $Na_2HPO_4$ | 1.2 g | 1.2 g | 1.2 g | 1.2 g |
| $KH_2PO_4$ | 0.6 g | 0.6 g | 0.6 g | 0.6 g |
| NaCl | 0.1 g | 0.1 g | 0.1 g | 0.1 g |
| $MgSO_4$ stock solution | 200 μl | 200 μl | 200 μl | 200 μl |
| $CaCl_2$ stock solution | 200 μl | 200 μl | 200 μl | 200 μl |
| Thiamine stock solution | 40 μl | 40 μl | 40 μl | 40 μl |
| $FeCl_3$ stock solution | 40 μl | 40 μl | 40 μl | 40 μl |
| Metal mixture stock solution | 100 μl | 100 μl | 100 μl | 100 μl |
| Total volume | 100 ml | 100 ml | 100 ml | 100 ml |

100% $D_2O$ was used for the preparation of media for selective protonation at side-chain methyl groups, whereas $H_2O$ was used for $^{13}C/^{15}N$ uniform labeling. Because TTHA1718 protein has no isoleucine residues, Cδ methyl-selective labeling at isoleucine residues from [U-$^{13}C$, 3,3-$^{2}H$] α-ketobutyrate is not included in this table.

into NMR tubes as an ~60% slurry with M9 medium containing 10% $D_2O$. We recommend monitoring the stability of *E. coli* samples by 2D $^{1}H$-$^{15}N$ HSQC spectra, followed by plating colony tests, which allows for a comparison between the initial and end-point health of the cells.

The present protocol does not require a specific set of NMR spectra. Any of the common 2D and 3D spectra used for the assignment of the polypeptide backbone and amino-acid side chains can be applied, provided that it yields useful information. **Table 2** summarizes the set of 2D and 3D NMR spectra that were used for the in-cell and *in vitro* structure determinations of TTHA1718. For all 3D NMR experiments, we use a nonlinear sampling scheme[19–21] for the indirectly observed dimensions to reduce experimental time. Briefly, about 1/8–1/4 of the points are chosen for measurement in a pseudorandom manner from the conventional regularly spaced grid of $t_1$,$t_2$ points (**Fig. 1**). The 2D maximum entropy method[26] is used for processing the nonlinearly sampled dimensions.

The 3D structures of proteins are obtained by structure calculation with the program CYANA[24] version 3.0 using automated NOE assignment[27] and torsion angle dynamics[28]. In addition to NOE distance restraints, backbone torsion angle restraints obtained from chemical shifts, e.g., using the program TALOS[29], can be added to the input, as well as distance restraints for hydrogen bonds in regular secondary structure elements. The 20 conformers with the lowest final CYANA target function values are embedded in a water shell and energy minimized against the AMBER force field[30] with the program OPALp[31,32] in the presence of the experimental restraints.

**TABLE 2** | NMR spectra recorded for the in-cell structure calculation of TTHA1718.

| For backbone assignment | For side-chain assignment | For distance restraint measurement |
|---|---|---|
| 2D $^{1}H$-$^{15}N$ HSQC | 2D $^{1}H$-$^{13}C$ HSQC | 3D $^{15}N$-separated |
| 3D HN(CO)CA | 2D $^{1}H$-$^{13}C$ HMQC | NOESY-HSQC |
| 3D HNCA | 3D | 3D $^{13}C$-separated |
| 3D HNCO | HBHA(CBCACO)NH | NOESY-HSQC |
| 3D HN(CA)CO | 3D H(CCCO)NH | 3D $^{13}C$/$^{13}C$-separated |
| 3D CBCA(CO)NH | 3D (H)C(CCO)NH | HMQC-NOESY-HMQC |
| 3D CBCANH | | (ALV-selective samples) |

## MATERIALS

### REAGENTS

- Expression plasmid encoding the protein of interest (for example, we used pET-11a (Novagen, cat. no. 69436-3) and pET-14b (Novagen, cat. no. 69660-3) expression systems for TTHA1718 and rat calmodulin, respectively)
- JM109 (DE3) *E. coli* (Promega, cat. no. P9801)
- BL21 (DE3) *E. coli* (Novagen, cat. no. 69450)
- $^{15}NH^{4}Cl$ (Spectra, cat. no. 5300)
- U-$^{13}$C-glucose (ProSpect Pharma, cat. no. PT100 803)
- [3-$^{13}$C] alanine (ISOTEC, cat. no. 489948)
- [U-$^{13}$C, 3-$^{2}$H] α-ketoisovalerate (CIL, cat. no. CDLM-4418-0)
- L-Lysine-$^{15}N_2$ HCl (ISOTEC, cat. no. 609021)
- L-Leucine (Sigma, cat. no. L8000)
- Glucose (Wako, cat. no. 041-00595)

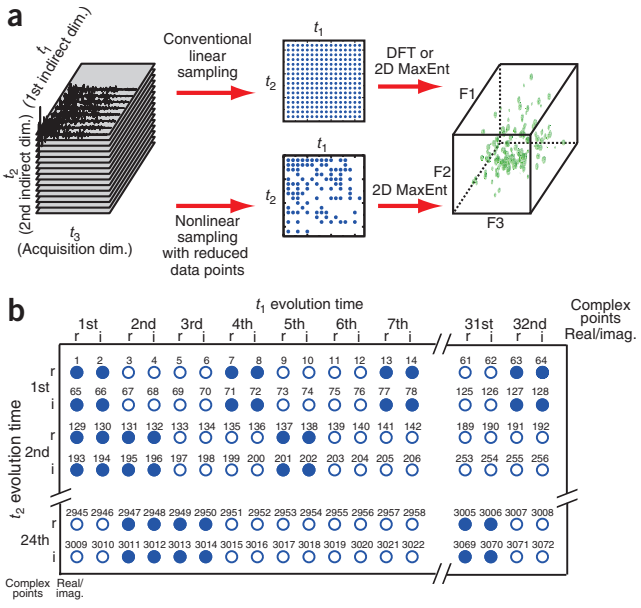


**Figure 1** | Rapid acquisition of 3D NMR spectra of proteins in living *E. coli* cells. (**a**) Schematic representation of rapid acquisition of 3D NMR spectra using a nonlinear sampling scheme. Adapted from reference 3. (**b**) An example of a nonlinear sampling scheme, in which data points are chosen for measurement in a pseudorandom manner from the regularly spaced grid of $t_1$,$t_2$ points (32 and 24 complex points, i.e., time domain data size 64 and 48, respectively). Real and imaginary data points are measured in pairs for both $t_1$ and $t_2$ evolution times. Chosen (filled blue circles) and omitted (blue circles) data points are represented with their AZARA-defined index numbers, which should be specified in the sampling lists for processing. (**c**) An example of a part of a script for 2D MaxEnt processing with AZARA. A brief explanation for each line is presented, following an exclamation mark. The parameter 'iter' defines the maximum number of iterations that the algorithm will use on one slice of data. Normally, the algorithm should converge in less iterations (typically within 10–20 iterations for 2D MaxEnt processing). The line starting with 'sample2' defines the sampling list (text format), in which AZARA-defined index numbers for chosen points are written in order of measurement. The usage of 'sample2' command is as follows: sample2 <d2 time domain data size> <d3 time domain data size> <sampling list name>. For further information, please see the AZARA manual pages (http://www.ccpn.ac.uk/azara/azara_docs/azara.html).

- $NH_4Cl$ (Wako, cat. no. 017-02995)
- Bacto tryptone (BD, cat. no. 211705)
- Bacto Yeast Extract (BD, cat. no. 212750)
- Bacto Agar (BD, cat. no. 214010)
- NaCl (Wako, cat. no. 191-01665)
- $Na_2HPO_4$ (Wako, cat. no. 197-02865)
- $KH_2PO_4$ (Wako, cat. no. 169-04245)
- $MgSO_4$ (Wako, cat. no. 137-12335)
- $CaCl_2$ (Wako, cat. no. 039-00475)
- $FeCl_3 \cdot 6H_2O$ (Wako, cat. no. 091-00872)
- $ZnSO_4 \cdot 7H_2O$ (Wako, cat. no. 264-00402)
- $CuSO_4 \cdot 5H_2O$ (Wako, cat. no. 039-04412)
- $MnSO_4 \cdot 5H_2O$ (Wako, cat. no. 130-13182)
- $H_3BO_3$ (Wako, cat. no. 021-02195)
- Thiamine hydrochloride (Wako, cat. no. 201-00852)
- Isopropyl thio-β-D-thiogalactoside (Wako, cat. no. 099-05013)
- $D_2O$ (Spectra, cat. no. 5150)
- Ampicillin sodium (Wako, cat. no. 018-10372)
- Bug Buster Protein Extraction Reagent (Novagen, cat. no. 70584-3)
- Benzamidine, hydrochloride (Calbiochem, cat. no. 199001)
- Simply Blue Safe Stain (Invitrogen, cat. no. LC6065)
- NuPAGE 12% Bis-TrisGel (Invitrogen, cat. no. NP0349BOX)
- NuPAGE MES SDS Running Buffer (20×; Invitrogen, cat. no. NP0002)
- NuPAGE LDS Sample Buffer (4×; Invitrogen, cat. no. NP0007)
- Mark 12 Unstained Standard (Invitrogen, cat. no. 354377)
- LB medium (10 $g l^{-1}$ Bacto tryptone, 5 $g l^{-1}$ Bacto Yeast Extract, 10 $g l^{-1}$ NaCl)
- $MgSO_4$ stock solution 1 M $MgSO_4$
- $CaCl_2$ stock solution 50 mM $CaCl_2$
- $FeCl_3$ stock solution 5 mM $FeCl_3 \cdot 6H_2O$
- Metal mixture stock solution (4 mM $ZnSO_4 \cdot 7H_2O$, 1 mM $MnSO_4 \cdot 5H_2O$, 4.7 mM $H_3BO_3$, 0.7 mM $CuSO_4 \cdot 5H_2O$)
- Thiamine stock solution 0.3 M thiamine hydrochloride
- M9 minimal medium for uniformly $^{13}C/^{15}N$-labeled samples (see **Table 1**)
- M9 minimal medium for selectively protonated side-chain methyl groups (see **Table 1** ).

**EQUIPMENT**

- NMR spectrometer ($^1H$ frequency 500 MHz or higher and equipped with a cryoprobe that is highly recommended)
- Linux computer system for data processing, analysis of spectra and structure calculations (multiple processors recommended)
- Software for processing nonlinearly sampled NMR data with the 2D maximum entropy method[26,33] (2D MaxEnt), e.g., AZARA 2.7 (W. Boucher).
- Software for interactive spectra analysis, e.g., ANSIG 3.3 (http://www.bio.cam.ac.uk/nmr/ccmr/public/ANSIG/ansig.html)[34,35], CcpNmr Analysis[36], NMRView (One Moon Scientific Inc.)[37,38], Sparky, XEASY (http://www.mol.biol.ethz.ch/groups/wuthrich_group/software)[39] and CARA (http://cara.nmr.ch/doku.php)
- CYANA 3.0 software package for automated NOESY assignment and structure calculation[24]
- TALOS[29] or TALOS + [40] software for deriving backbone torsion angle restraints on the basis of the chemical shifts

**EQUIPMENT SETUP**

**NMR spectrometer** The NMR spectrometer must be equipped for triple-resonance ($^1H/^{13}C/^{15}N$) experiments with pulsed field gradients. As sensitivity is one of the limiting factors of in-cell NMR experiments, a cryogenic probe head and a magnetic field strength corresponding to a $^1H$ frequency of at least 500 MHz are highly recommended.

**NMR pulse sequences** These should be prepared so as to control sampling points according to sampling schedule lists as opposed to conventional sampling of every point on a regularly spaced grid. For our NMR spectrometer (Bruker Avance 600), pulse sequences were modified according to the procedure reported by Rovnyak *et al.*[21], in which each indirect point to be acquired from the conventional regularly spaced grid of $t_1$,$t_2$ points is defined by 2–4 lines (digits) in variable counter (VC) lists. Pulse sequences are available from the corresponding author at http://www.comp.tmu.ac.jp/osbc/GROUP/ITO/ito_nls_utility.html. There are programs available to generate VC lists for nonlinear sampling: e.g., COAST[21] (http://gwagner.med.harvard.edu/) and NUSSAMPLER[41] (http://groups.google.com/group/mddnmr/), in which sampling points are selected using a weighting function for each dimension according to the type of time evolution: exponential (or other decaying) functions for conventional evolution dimensions; constant functions for constant-time evolution dimensions. In our group, a simple computer program was used to generate VC lists for nonlinear sampling, which is also at http://www.comp.tmu.ac.jp/osbc/GROUP/ITO/ito_nls_utility.html.

**Computer and software setup** Structure calculations with the program CYANA[24] and restrained energy refinements with the program OPALp[31,32] can be run efficiently on Linux cluster systems. The AZARA 2.7 software can be used for processing nonlinearly sampled data with 2D MaxEnt[26]. The sampling lists for data processing with the AZARA 2.7 software should be generated from the corresponding VC lists for the NMR measurement. The Rowland NMR Toolkit (http://webmac.rowland.org/rnmrtk/)[33] is an alternative software for this type of processing. Note that, in addition to MaxEnt, other processing procedures, such as multidimensional decomposition[42] and multidimensional Fourier transform[43,44], have been used for processing nonlinearly sampled data. An interactive NMR spectrum analysis program is needed for obtaining the resonance assignment from the spectra. In our laboratory, either an OpenGL version of ANSIG 3.3 software[34,35] (http://www.bi.a.u-tokyo.ac.jp/~takeshi/ansig4opengl/) or CcpNmr Analysis software[36] were used. Alternative programs such as NMRView (One Moon Scientific Inc.)[37,38], Sparky, XEASY[39] and CARA can also handle AZARA-processed NMR spectra after format conversion.

## PROCEDURE

### Preparation of proteins inside living *E. coli* cells

**1|** Transform *E. coli* cells with the overexpression plasmid.

**2|** Grow *E. coli* cells in 2 ml LB media at 37 °C with shaking to a high $OD^{600}$ of ~2.0.

● **TIMING** ~10 h

**3|** Subculture *E. coli* cells (100 µl) in 100 ml unlabeled M9 media, and incubate the culture at 37 °C until the $OD^{600}$ reaches 0.5–0.6.

● **TIMING** 12–14 h

**4|** Centrifuge the culture at ~800 *g* for 20 min at room temperature (25 °C).

**5|** Decant the supernatant. Centrifuge the pellet again at ~800*g* for 5 min at room temperature.

**6|** Remove the supernatant by pipetting, and resuspend the cells in 100 ml of stable isotope-labeled M9 media.

**!** **CAUTION** The samples selectively labeled at side-chain methyl groups are expressed in 100% $D_2O$.

**7|** Incubate the cells at 37 °C without shaking for 1 h.

## BOX 1 | CHECKING THE SAMPLE CONDITION, THE VIABILITY OF CELLS, AND THE LOCALIZATION OF OVEREXPRESSED PROTEINS DURING IN-CELL NMR EXPERIMENTS

1. To ensure that the observed NMR spectrum represents intracellular protein and that the signals are not caused by proteins released from the bacteria because of cell lysis, after each NMR measurement, transfer a small amount (~10 µl) from the NMR sample to a microfuge tube and centrifuge until all bacteria are collected in a pellet. Store the pellet and supernatant separately for SDS–PAGE.
2. The viability of cells during NMR experiments can be investigated by a plating colony test. Spread a small volume of the NMR sample (~10 µl) taken before and after the experiments on LB plates containing the appropriate antibiotic. Incubate the plates overnight at 37 °C and count the colonies. We experienced that, under our measurement condition, viability decreased slowly for at least 6 h, at which time point viability was 85 ± 11%, then decreased with gradually increasing pace. We recommend plotting the viability curve for new samples before designing the measurement schedule of in-cell NMR spectra.
3. Analyze the localization of overexpressed proteins by measuring 2D $^{1}H$-$^{15}N$ HSQC spectra of spheroplasts and periplasmic extracts, which are fractionated from target protein-expressing $^{15}N$-labeled cells by lysozyme-EDTA treatment using the conditions described by Thorstenson *et al.*[46]. Ensure spheroplast formation by light microscopy. The localization of the protein in *E. coli* cells can also be predicted from its amino-acid sequence by PSORTb version 2.0 (refs. 47,48) (http://www.psort.org/psortb/) and SignalP 3.0 (refs. 49,50) (http://www.cbs.dtu.dk/services/SignalP/).
4. Optionally, prepare purified protein for *in vitro* NMR experiments to compare the effects of in-cell and *in vitro* conditions on the structure and physical properties.

**8|** Induce the production of the target protein (for example, by adding isopropyl thio-β-D-thiogalactoside to a final concentration of 0.5 mM).

**9|** Continue protein expression with shaking at optimal temperature.
● **TIMING** ~3 h

**10|** Harvest the cells by centrifugation at ~400*g* for 30 min at room temperature.

**11|** Remove the supernatant by aspiration, and resuspend the cells by adding small amounts of unlabeled M9 media (140–160 µl) and carefully pipetting the solution up and down until the entire cell pellet has been suspended.

**12|** Add $D_2O$ (10% of the final sample volume) to the bacterial slurry, and transfer the sample to an NMR tube. The condition and viability of the sample should be checked (**Box 1**).
▲ **CRITICAL STEP** Ensure that the proteins providing NMR spectra are indeed inside the living cells, and that the contribution from extracellular proteins is negligible. The concentration of the expressed protein in *E. coli* cells can be estimated by comparing the density of the Coomassie-stained bands in SDS–PAGE with those of proteins with similar molecular size and known concentration. The localization of overexpressed proteins can be predicted from its amino-acid sequence, and can also be checked experimentally (**Box 1**).
**? TROUBLESHOOTING**

**NMR measurements and data processing**

**13|** Before the measurement of the in-cell NMR sample, shim the magnetic field with a separate NMR sample containing unlabeled M9 media (10% $D_2O$), which is prepared with the same sample length as for the in-cell NMR sample.

**14|** Insert the in-cell NMR sample into the magnet, and tune the probe head.

**15|** Check the sample's condition by measuring 1D $^{1}H$-NMR spectra and 1D (or 2D) $^{1}H$-$^{15}N$ HSQC spectra.

**16|** Collect 3D NMR spectra required for chemical shift assignments and structure calculation (see **Table 2**). For all 3D NMR experiments, a nonlinear sampling scheme[19–21] should be used for the indirectly observed dimensions to reduce measurement time. Although optimal reduction should be considered in each case, approximately 1/4–1/8 of the data points were typically selected in a pseudorandom manner from the conventional regularly spaced grid of $t_1$,$t_2$ points in our case. Prepare a fresh sample for each 3D experiment.
● **TIMING** The duration of the NMR experiments for an in-cell NMR sample must be set considering the lifetime of the *E. coli* cells under measurement conditions.
▲ **CRITICAL STEP** With the nonlinear sampling scheme, the duration of each 3D experiment is reduced to 2–3 h. Repeat the measurement of each 3D experiment several (3–4) times interleaved with a short 2D $^{1}H$-$^{15}N$ HSQC experiment used to monitor the condition of the sample. Combine these 3D data to generate a new data set with improved signal-to-noise ratio up to the point that the 2D spectra show significant changes. In the case of TTHA1718, two 3D data sets were combined.

**17|** Process all 3D spectra using the 2D maximum entropy method (**Box 2**).
**? TROUBLESHOOTING**

**Spectra analysis and structure determination**

**18|** Perform backbone and side-chain resonance assignments by analyzing the 3D NMR spectra listed in **Table 2**.
▲ **CRITICAL STEP** Backbone and side-chain resonances should be assigned as completely as possible to assist the automated NOE assignment in the subsequent structure calculation.

**19|** Perform side-chain methyl resonance assignments by analyzing 3D (H)CC(CO)NH and H(CCCO)NH spectra measured on selectively methyl-protonated samples. Intraresidue and sequential NOEs involving methyl protons are also used for the assignment. $^{1}$H-$^{13}$C HMQC spectra of in-cell NMR samples with different methyl-selective labeling patterns are used for amino-acid classification of methyl $^{1}$H-$^{13}$C correlation cross-peaks.

**20|** Analyze 3D $^{15}$N-separated NOESY-HSQC and 3D $^{13}$C-separated NOESY-HSQC spectra measured on the uniformly labeled sample. For the analysis of NOEs involving methyl protons, 3D $^{13}$C-separated NOESY-HSQC and 3D $^{13}$C/$^{13}$C-separated HMQC-NOE-HMQC spectra measured on methyl-selectively protonated samples are useful.
**! CAUTION** In contrast to $^{15}$N-labeling, uniform $^{13}$C-labeling gives rise to a considerable number of 'background' cross-peaks (**Fig. 2**)[23]. NOE cross-peaks in the 3D $^{13}$C-separated NOESY should be analyzed carefully and selected only if they are highly likely to correlate resonances of the target protein.

**21|** Prepare NOESY peak lists containing the chemical shifts and volumes (or intensities) of the cross-peaks in the NOESY spectra.
**! CAUTION** Use consistent chemical shift referencing in NOESY spectra and for the chemical shifts determined in Steps 22 and 23.

**22|** Prepare backbone torsion angle restraints on the basis of chemical shifts with the programs TALOS[29] or TALOS+[40].

**23|** Prepare distance restraints for hydrogen bonds for those positions in the regular secondary structure regions in which the existence of hydrogen bonds is strongly supported by medium-range or interstrand NOEs.

## BOX 2 | PROCESSING OF NONLINEARLY SAMPLED 3D NMR DATA WITH THE 2D MAXIMUM ENTROPY METHOD IMPLEMENTED IN THE AZARA SOFTWARE

1. Prepare a sampling list file (text format) for 2D MaxEnt processing. Examples of all files and scripts described in this section are made available at the URL, http://www.comp.tmu.ac.jp/osbc/GROUP/ITO/ito_nls_utility.html. The list describes the points that have been sampled in the input data set relative to the usual uniformly sampled set. Each integer in the list identifies the location of an element of the hypercomplex points sampled in the conventional sampling scheme (see **Fig. 1b**). Ensure that the total number of points defined in the sampling list file is consistent with the number of collected FIDs.
2. Generate a parameter file that is used to describe the dimensions, referencing, etc., of the data set. Ensure that the product of the number of real points for indirect dimensions (d2 and d3 dimensions), for which nonlinear sampling is applied, is equal to the total number of collected FIDs.
3. Generate a script file for the FFT processing of the acquisition dimension (d1), in which commands for phase correction should be omitted. Process the data with the script.
4. Determine the 0th and 1st order phase correction parameters for the dimension d1 using the phase utility of 'plot2'.
5. Generate a script file consisting of the FFT part for the dimension d1 (including phasing with the phase parameters determined above) and the 2D MaxEnt part for the d2 and d3 dimensions (**Fig. 1c**). The most important parameter for MaxEnt processing with Azara is the noise level, which is usually estimated by using the equation:

$$\text{noise} = \frac{\text{standard deviation of data values for noise regions}}{\sqrt{(\text{number of complex points for d2} \times \text{number of complex points for d3})}}$$

However, it is impossible to use this equation for nonlinearly sampled data, as FFT-processed spectra cannot be produced. Instead, the noise level has to be estimated from spectra with similar acquisition parameters, and modified by trial and error with monitoring of the convergence of the calculations (typically within 10–20 iterations).
6. Perform processing of the data with the script file. It is recommended that the MaxEnt algorithm be tested on a slice of data before it is run on the entire data file to ensure that convergence occurs for the chosen parameters.
7. Check the calculated spectra with 'plot2'.

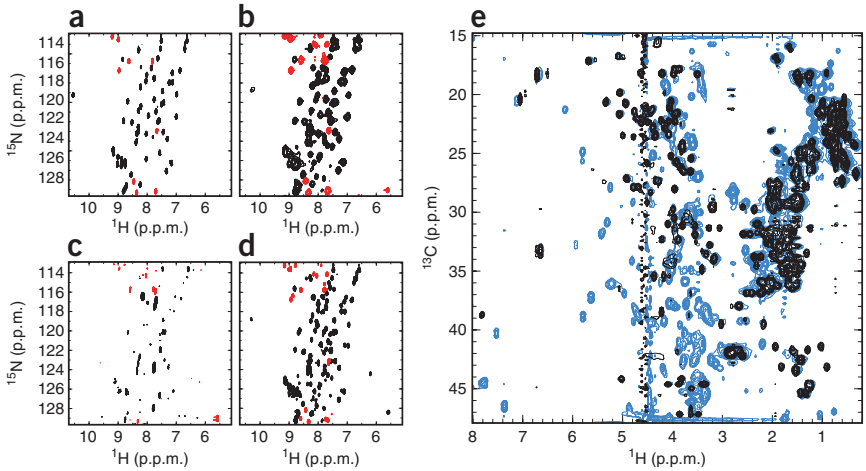


**Figure 2** | 2D $^{1}$H-$^{15}$N or $^{1}$H-$^{13}$C HSQC spectra of TTHA1718 in living *E. coli* cells. (**a**) The $^{1}$H-$^{15}$N HSQC spectrum of purified TTHA1718. (**b**) The $^{1}$H-$^{15}$N HSQC spectrum of a TTHA1718 in-cell NMR sample immediately after sample preparation. (**c**) The $^{1}$H-$^{15}$N HSQC spectrum of the supernatant of the in-cell sample after 6 h measurement. (**d**) The $^{1}$H-$^{15}$N HSQC spectrum of the lysate of the harvested cells after 6-h NMR measurement. For **a**–**d**, negative contours originating from aliased cross-peaks are represented in red. (**e**) Overlay of the $^{1}$H-$^{13}$C HSQC spectra of purified TTHA1718 (black) and *E. coli* cells expressing TTHA1718 (blue). Adapted from reference 3.

**24|** Use the program CYANA[24] to run multiple (typically seven) iterative cycles of automated NOE assignment[27] and structure calculation by simulated annealing using torsion angle dynamics[28] (**Table 3**)[45].

● **TIMING** 0.2–5 h, depending on computer performance and protein size.

■ **PAUSE POINT** CYANA calculations can usually be accomplished in <1 h of unattended computation time, but may take several hours depending on the performance of the computer, the size of the protein, the quantity of NMR data and the number of structures computed in CYANA.

**!** **CAUTION** The reliability of the structure obtained by automated NOE assignment with CYANA should be ascertained by an RMSD of less than 3 Å for the backbone atoms, excluding flexible regions, in the first cycle of the structure calculation[3]. If necessary, manually assigned NOE peaks can also be included in the CYANA calculation.

**TABLE 3** | Parameters for in-cell structure calculations with CYANA.

| Parameter | Typical value |
|---|---|
| NOE assignment and structure calculation cycles | 7 |
| Tolerance for $^{1}$H/$^{13}$C/$^{15}$N chemical shift matching | 0.03/0.3/ 0.3 p.p.m. |
| Median distance for automated NOE calibration | 4.2 Å |
| Number of conformers calculated | 100 |
| Number of conformers analyzed | 20 |
| Number of torsion angle dynamics steps per conformer | 10,000 |

**25|** Analyze the results of the CYANA calculation. Check the extent of NOESY cross-peak assignments, the summary table of the structure calculation provided by CYANA and the final 3D structure with a molecular graphics program.

**? TROUBLESHOOTING**

**26|** Perform energy refinement of the structure against the AMBER force field[30].

**? TROUBLESHOOTING**

Troubleshooting advice can be found in **Table 4**.

**TABLE 4** | Troubleshooting table.

| Step | Problem | Possible reason | Solution |
|---|---|---|---|
| 12 | No, or very weak cross-peaks from the target protein in in-cell NMR spectra | Unknown. Motional restrictions due to nonspecific interactions inside *E. coli* cells can be one of the possible reasons | Usually drastic improvement cannot be guaranteed, but it is worth altering the conditions for protein expression, e.g., host *E. coli* strains, timing of induction, optimal temperature and incubation times |
| | Leakage of the target protein from cells during NMR experiments | Unknown | As above. In one case, leakage was largely prevented by reducing the temperature for protein expression and NMR measurement<br>Li *et al.*[51] reported that encapsulation in alginate microcapsules stabilizes *E. coli* cells and prevents leakage |
| 17 | No, or very slow convergence in 2D MaxEnt processing | Noise level is too low or rate is too low | Increase the value of 'noise' parameter in the script. Check the convergence of the calculations (within 10–20 iterations) indicated in the output log files or increase rate |

(Continued)

**TABLE 4** | Troubleshooting table (continued).

| Step | Problem | Possible reason | Solution |
|---|---|---|---|
| | Excessively rapid convergence or oscillating ω value in 2D MaxEnt processing | Noise level is too high or rate is too high | Decrease the value of the 'noise' or 'rate' parameter |
| 25 | No convergence in structure calculation | Too few, or inconsistent NOE distance restraints | Check chemical shift assignment, NOESY peak picking, chemical shift referencing, and the specification of the spectral dimensions in the peak list headers |



## ANTICIPATED RESULTS

The protocol was applied to a model system, the *T. thermophilus* HB8 TTHA1718 gene product, a putative heavy metal-binding protein of 66 amino acids that was overexpressed in *E. coli* to a concentration of 3–4 mM (ref. 3).

The $^{1}H$-$^{15}N$ HSQC spectra of TTHA1718 *in vitro* and in-cell are presented in **Figure 2a,b**. The in-cell spectrum shows a much broader line shape for both $^{1}H$ and $^{15}N$ dimensions. The virtual identity of the $^{1}H$-$^{15}N$ HSQC spectra recorded immediately after sample preparation (**Fig. 2b**) and after 6 h in an NMR tube at 37 °C (data not shown) shows 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 NMR spectra are indeed inside the living cells, and that the contribution from extracellular proteins is negligible. Most $^{1}H$-$^{15}N$ HSQC cross-peaks disappeared on removal of bacteria by gentle centrifugation (**Fig. 2c**), whereas the spectrum of the lysate of the harvested cells showed much sharper cross-peaks (**Fig. 2d**). **Figure 2e** shows $^{1}H$-$^{13}C$ HSQC spectra of TTHA1718 *in vitro* (black) and in-cell (blue). In contrast to $^{15}N$-labeling, uniform $^{13}C$-labeling gave rise to a considerable number of 'background' cross-peaks[23].

**Figure 3** shows $^{13}C(F1)$-$^{1}H^{N}(F3)$ or $^{1}H(F1)$-$^{1}H^{N}(F3)$ slices corresponding to the amide $^{15}N$ chemical shift of Lys30 (120.06 p.p.m.) from 3D triple-resonance NMR spectra for the backbone and side-chain resonance assignment measured for TTHA1718 in living *E. coli* cells.

**Figure 4** shows slices from 3D NOESY-type experiments for TTHA1718 in living *E. coli* cells. The 3D $^{15}N$-separated NOESY-HSQC spectrum (**Fig. 4a**) and the 3D $^{13}C$-separated NOESY-HSQC spectrum (**Fig. 4b**) were measured on uniformly labeled TTHA1718 in living *E. coli* cells. The 3D $^{13}C/^{13}C$-separated HMQC–NOE–HMQC spectrum (**Fig. 4c**) was measured on a TTHA1718 in-cell NMR sample, in which methyl groups of Ala, Leu and Val residues were selectively $^{1}H/^{13}C$ labeled.

**Figure 5a** shows the 3D structures of TTHA1718 in *E. coli* cells calculated using the program CYANA[24] on the basis of NOE distance restraints, backbone torsion angle restraints and restraints for hydrogen bonds. The resulting structure is well converged with a backbone RMSD of 0.96 Å to the mean coordinates, and is similar to the structure that was determined independently *in vitro* from a purified sample (**Fig. 5b**). The backbone RMSD between the in-cell and *in vitro* structures is 1.16 Å (**Fig. 5c**). 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. Particularly, in the putative heavy metal-binding loop, interactions with metal ions in the *E. coli* cytosol may affect the conformation of the region. Long-range methyl–methyl NOEs were indispensable for the precise determination of 3D structures, as the convergence of structures dropped drastically to a backbone RMSD of 5.46 Å, when calculated without distance restraints derived from NOEs involving methyl groups obtained in selectively methyl-protonated in-cell NMR samples (**Fig. 5d**).

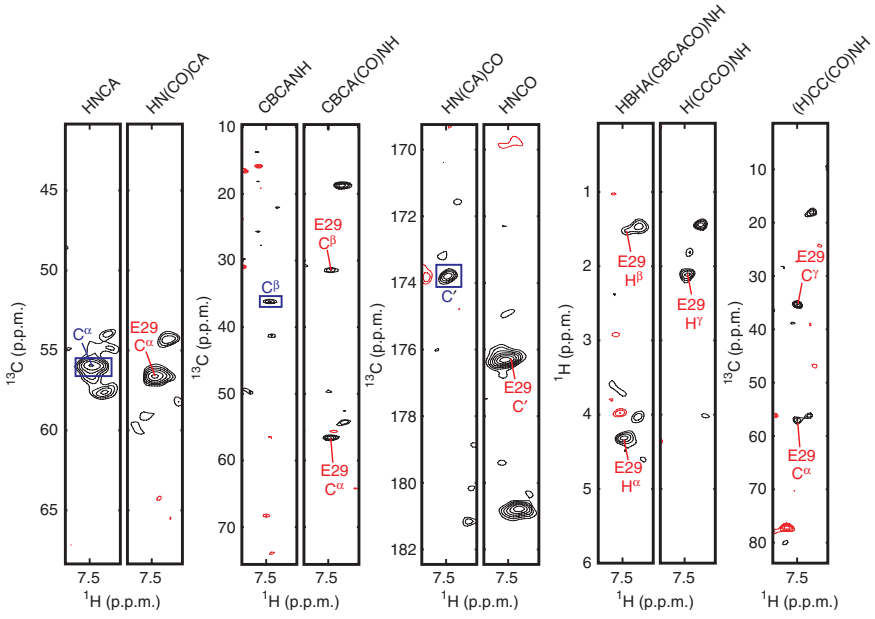


**Figure 3** | $^{13}C(F1)$-$^{1}H^{N}(F3)$ or $^{1}H(F1)$-$^{1}H^{N}(F3)$ cross-sections corresponding to the amide $^{15}N$ chemical shift of Lys30 (120.06 p.p.m.) from 3D HNCA, HN(CO)CA, CBCANH, CBCA(CO)NH, HN(CA)CO, HNCO, HBHA(CBCACO)NH, H(CCCO)NH and (H)CC(CO)NH spectra (black) used for backbone and side-chain resonance assignments of TTHA1718 in living *E. coli* cells. The cross-peaks due to sequential correlations are assigned in red. Intraresidue correlations are indicated by blue boxes and annotated.

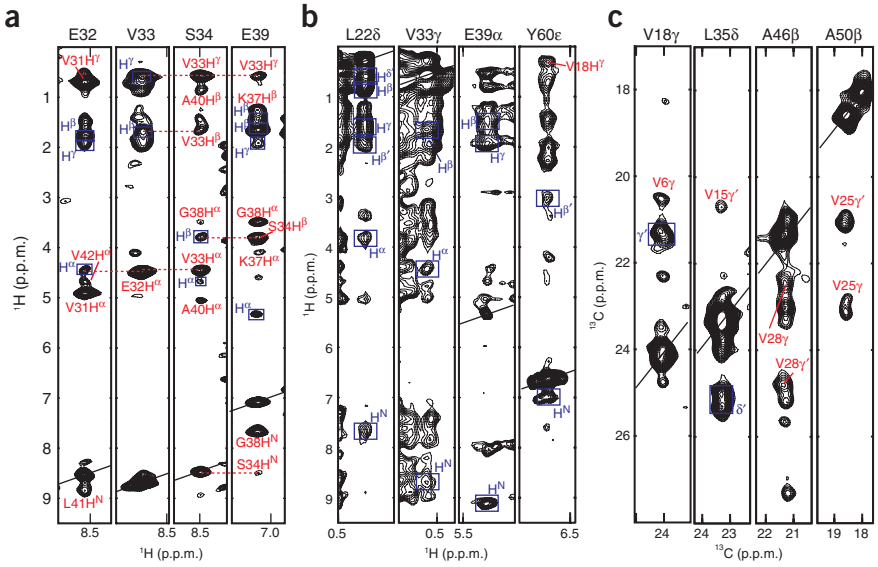


**Figure 4** | Collection of nuclear Overhauser effect-derived distance restraints in TTHA1718 in living *E. coli* cells. (**a**) $^{1}H(F1)$-$^{1}H^{N}(F3)$ cross-sections corresponding to the $^{15}N$ chemical shifts of selected backbone amide groups extracted from the 3D $^{15}N$-separated NOESY-HSQC spectrum. The cross-peaks due to interresidual NOEs are assigned in red. Intraresidual NOEs are indicated by blue boxes and annotated. (**b**) $^{1}H(F1)$-$^{1}H(F3)$ cross-sections corresponding to the $^{13}C$ chemical shifts of representative carbon nuclei extracted from the 3D $^{13}C$-separated NOESY-HSQC spectrum. The inter- and intraresidue NOEs are indicated as in **a**. (**c**) $^{13}C(F1)$-$^{13}C(F2)$ cross-sections corresponding to the $^{13}C$ frequencies of representative methyl groups extracted from the 3D $^{13}C/^{13}C$-separated HMQC–NOE–HMQC spectrum. The inter- and intraresidue NOEs are indicated as in **a**. Adapted from reference 3.

In conclusion, in-cell protein structure determination by NMR opens new avenues for studying in atomic resolution, the manner in which protein conformations change in response to biological events in living environments. The approach provides tools for investigating in living cells the effects of molecular crowding in the cytosol, protein stability and covalent protein modification, the conformations of proteins that are intrinsically disordered *in vitro* and the 3D structures of proteins that are otherwise unstable and difficult to purify.

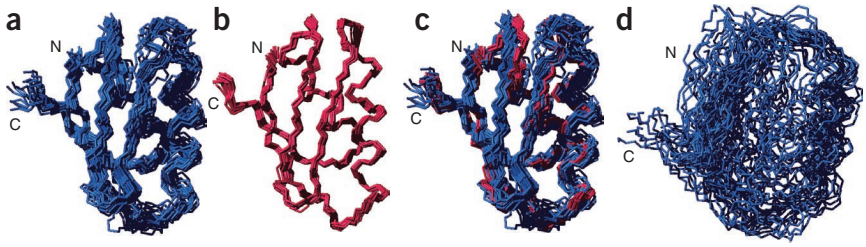


**Figure 5** | NMR solution structure of TTHA1718 in living *E. coli* cells. (**a**) A superposition of the 20 final structures of TTHA1718 in living *E. coli* cells, showing backbone (N, $C^{\alpha}$, C′) atoms. (**b**) A superposition of the 20 final structures of purified TTHA1718 *in vitro*. (**c**) A comparison of TTHA1718 structures in living *E. coli* cells and *in vitro*. The best-fit superposition of backbone (N, $C^{\alpha}$, C′) atoms of the two conformational ensembles is shown with the same color code in **a** and **b**. (**d**) A superposition of the 20 final structures of TTHA1718 in living *E. coli* cells calculated without distance restraints derived from NOEs involving methyl groups obtained in methyl-selectively protonated in-cell NMR samples. Adapted from reference 3.

**ACKNOWLEDGMENTS** We thank Professor Seiki Kuramitsu for providing the plasmid encoding TTHA1718. This work was supported in part by the CREST program of the Japan Science and Technology Agency (JST), the Molecular Ensemble Program of 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 article. T.I. developed the protocol for structure calculation and refinement. A.S., D.S., J.H., T.H., M.Y., N.H. and T.M. developed the protocols, including sample preparation and characterization, data acquisition and resonance assignment. Y.S., M.M., D.N. and M.W. contributed the protocol development for rapid NMR data acquisition and MaxEnt processing.

**COMPETING FINANCIAL INTERESTS** The authors declare no competing financial interests.

Published online at http://www.natureprotocols.com/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Ellis, R.J. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* **26**, 597–604 (2001).
2. Pielak, G.J. *et al.* Protein nuclear magnetic resonance under physiological conditions. *Biochemistry* **48**, 226–234 (2009).
3. Sakakibara, D. *et al.* Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **458**, 102–105 (2009).
4. Serber, Z. & Dotsch, V. In-cell NMR spectroscopy. *Biochemistry* **40**, 14317–14323 (2001).
5. Serber, Z., Corsini, L., Durst, F. & Dötsch, V. In-cell NMR spectroscopy. *Method. Enzymol.* **394**, 17–41 (2005).
6. Serber, Z. *et al.* Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat. Protoc.* **1**, 2701–2709 (2006).
7. Burz, D.S., Dutta, K., Cowburn, D. & Shekhtman, A. Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat. Methods* **3**, 91–93 (2006).
8. Burz, D.S., Dutta, K., Cowburn, D. & Shekhtman, A. In-cell NMR for protein-protein interactions (STINT-NMR). *Nat. Protoc.* **1**, 146–152 (2006).
9. Reckel, S., Hänsel, R., Löhr, F. & Dötsch, V. In-cell NMR spectroscopy. *Prog. Nucl. Mag. Res. Sp.* **51**, 91–101 (2007).
10. Selenko, P. & Wagner, G. Looking into live cells with in-cell NMR spectroscopy. *J. Struct. Biol.* **158**, 244–253 (2007).

11. Serber, Z. *et al.* High-resolution macromolecular NMR spectroscopy inside living cells. *J. Am. Chem. Soc.* **123**, 2446–2447 (2001).
12. 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).
13. Wieruszeski, J.M., Bohin, A., Bohin, J.P. & Lippens, G. *In vivo* detection of the cyclic osmoregulated periplasmic glucan of *Ralstonia solanacearum* by high-resolution magic angle spinning NMR. *J. Magn. Reson.* **151**, 118–123 (2001).
14. 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).
15. Sakai, T. *et al.* In-cell NMR spectroscopy of proteins inside *Xenopus laevis* oocytes. *J. Biomol. NMR* **36**, 179–188 (2006).
16. Bodart, J.F. *et al.* NMR observation of Tau in *Xenopus* oocytes. *J. Magn. Reson.* **192**, 252–257 (2008).
17. Inomata, K. *et al.* High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **458**, 106–109 (2009).
18. Ogino, S. *et al.* Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. *J. Am. Chem. Soc.* **131**, 10834–10835 (2009).
19. 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).
20. 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).
21. 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).
22. Rosen, M.K. *et al.* Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol.* **263**, 627–636 (1996).
23. 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).
24. Güntert, P. Automated NMR protein structure calculation. *Prog. Nucl. Mag. Res. Sp.* **43**, 105–125 (2003).
25. Williams, S.P., Haggie, P.M. & Brindle, K.M. $^{19}$F NMR measurements of the rotational mobility of proteins *in vivo*. *Biophys. J.* **72**, 490–498 (1997).
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. 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).
29. 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).
30. 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).
31. Luginbühl, P., Güntert, P., Billeter, M. & Wüthrich, K. The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. *J. Biomol. NMR* **8**, 136–146 (1996).
32. Koradi, R., Billeter, M. & Güntert, P. Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun.* **124**, 139–147 (2000).
33. Hoch, J.C. & Stern, A.S. *NMR Data Processing* (Wiley-Liss, New York, 1996).
34. 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).
35. 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).
36. Vranken, W.F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
37. Johnson, B.A. & Blevins, R.A. NMR view: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
38. Johnson, B.A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352 (2004).
39. Bartels, C., Xia, T., Billeter, M., Güntert, P. & Wüthrich, K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **6**, 1–10 (1995).
40. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
41. Jaravine, V.A. & Orekhov, V.Y. Targeted acquisition for real-time NMR spectroscopy. *J. Am. Chem. Soc.* **128**, 13421–13426 (2006).
42. Jaravine, V., Ibraghimov, I. & Orekhov, V.Y. Removal of a time barrier for high-resolution multidimensional NMR spectroscopy. *Nat. Methods* **3**, 605–607 (2006).
43. Marion, D. Fast acquisition of NMR spectra using Fourier transform of non-equispaced data. *J. Biomol. NMR* **32**, 141–150 (2005).
44. Kazimierczuk, K., Zawadzka, A., Kozminski, W. & Zhukov, I. Random sampling of evolution time space and Fourier transform processing. *J. Biomol. NMR* **36**, 157–168 (2006).
45. Takeda, M., Ikeya, T., Güntert, P. & Kainosho, M. Automated structure determination of proteins with the SAIL-FLYA NMR method. *Nat. Protoc.* **2**, 2896–2902 (2007).
46. 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).
47. Gardy, J.L. *et al.* PSORT-B: improving protein subcellular localization prediction for Gram-negative bacteria. *Nucleic Acids Res.* **31**, 3613–3617 (2003).
48. 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).
49. 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).
50. 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).
51. 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).