High-Resolution Protein 3D Structure Determination in Living Eukaryotic Cells

Takashi Tanaka¹, Teppei Ikeya*, Hajime Kamoshida, Yusuke Suemoto, Masaki Mishima, Masahiro Shirakawa, Peter Güntert, and Yutaka Ito*

Abstract: Proteins in living cells interact specifically or nonspecifically with an enormous number of biomolecules. To understand the behavior of proteins under intracellular crowding conditions, it is indispensable to observe their three-dimensional (3D) structures at the atomic level in a physiologically natural environment. We demonstrate the first de novo protein structure determinations in eukaryotes with the sf9 cell/baculovirus system using NMR data from living cells exclusively. The method was applied to five proteins, rat calmodulin, human HRas, human ubiquitin, T. thermophilus HB8 TTHA1718, and Streptococcus protein GB 1 domain. In all cases, we could obtain structural information from well-resolved in-cell 3D nuclear Overhauser effect spectroscopy (NOE) data, suggesting that our method can be a standard tool for protein structure determinations in living eukaryotic cells. For three proteins, we achieved well-converged 3D structures. Among these, the in-cell structure of protein GB 1 domain was most accurately determined, demonstrating that a helix-loop region is tilted away from a β-sheet compared to the conformation in diluted solution.

Biomacromolecules occupy a significant fraction of the intracellular volume (resulting in molecular crowding) in which proteins are exposed to the excluded-volume effect, specific and non-specific interactions, and various dynamic intracellular processes. Their biophysical properties under these effects, particularly their molecular structures at the atomic level, are not fully understood. Therefore, it is indispensable to elucidate their native structures and dynamics in the physiologically natural environment inside cells, and to determine whether there are differences in the three-dimensional (3D) structures of the biomacromolecules in cells compared to their diluted solution state. In-cell NMR is currently the only tool with which to observe proteins and deoxyribonucleic acid (DNA) at atomic resolution in living biological systems. It also provides direct observations of protein behaviors in conjunction with chemical compounds that are potential targets for drug screening inside cells.

Although in-cell NMR studies in various eukaryotic cells have become possible by either expressing target proteins inside cells or by introducing stable isotope-enriched proteins from outside, high-resolution protein 3D structures have been determined only in Escherichia coli cells. To date, the achievable target-protein concentration in eukaryotic cells was too low to obtain a sufficient number of nuclear Overhauser effect (NOE)-derived distance restraints. In the meantime, in-cell NMR studies of human-cultured cells have revealed that the intracellular environment does indeed influence the protein folding stability and reduces the volume occupied by intrinsically disordered proteins.

Recently, protein global folds in cells were obtained by exploiting NMR chemical shifts and paramagnetic NMR effects induced by intracellularly stable lanthanoid-binding tags. The 3D protein structure prediction software Rosetta was used. 3D structures of the Streptococcus protein GB 1 domain inside Xenopus oocytes were deduced. However, the structures did not yield sufficiently detailed side-chain conformations, which are expected to be predominantly affected by the intracellular environment and essential for the function of proteins and applications such as drug discovery. To elucidating the subtle difference between the structures in vitro and in eukaryotic cells, it remains necessary to achieve de novo 3D protein structure determination from NOE-derived distance restraints between side-chains. Therefore, we improved the procedure that had been utilized for proteins at approximately a concentration of 250 μM in E. coli cells. Herein we show the first de novo protein structure determinations with high accuracy and precision in living eukaryotic cells. The method uses the sf9 cell/baculovirus system and the structure determinations were achieved exclusively based on information from 3D heteronuclear multidimensional NMR spectra and NOE-derived distance restraints.

As model systems, three small- and two medium-sized proteins were selected: Streptococcus protein GB 1 domain...
(57 a.a., 7 kDa; henceforth referred to as GB1), the *T. thermophilus* HB8 TTHA1718 gene product (66 a.a., 7 kDa), human ubiquitin with the three alanine mutations L8A, I44A, and V70A (74 a.a., 8 kDa; Ub3A), rat calmodulin (148 a.a., 17 kDa; CaM), and C-terminally truncated human HRas (residues 1–171, 19 kDa; HRas). The experimental scheme is presented in Scheme S1 (Supporting Information). Baculoviruses for expressing these proteins were constructed using the Bac-to-Bac system (Invitrogen). The concentration of GB1 in the sf9 cells was predicted to approximately 129 µM (Supporting Information, Figure S1). Considering that the maximal natural concentration of a protein in bacterial and human cells is a few dozen to hundreds of µM, our GB1 in sf9 NMR samples mimic the conditions of a physiologically natural environment well. In our previous work, approximately 80% of the backbone NMR resonances of GB1 were assigned exclusively from 3D triple-resonance NMR spectra in sf9 cells,[5a] while 3D spectra for side-chain resonance assignment and 3D nuclear Overhauser effect spectroscopy (NOESY) spectra suffered from the short lifetime of the cells. The cell viability fell below 90% after approximately 8 hours of measurements at 28°C in an NMR sample tube (Supporting Information, Figure S2). Therefore, we introduced a bioreactor system that supplies fresh medium into the NMR tube continuously.[11] This strategy prolonged the lifetime of the cells in the NMR tube to be almost comparable to that under “optimum” condition in culture flasks, maintaining >90% cell viability as well as protein stability in the cells for at least 24 hours (Supporting Information, Figures S2–4). Moreover, the bioreactor is also effective for removing extracellular proteins (Supporting Information, Figure S3), thus guaranteeing that only proteins inside sf9 cells contribute to the in-cell NMR spectra. All 3D NMR data were sparsely sampled[12] and reconstructed by quantitative maximum entropy (QME),[5b] which played a crucial role in producing improved sensitivity of the in-cell spectra and achieving their exact analysis. The 3D NOESY spectra of GB1, Ub3A, and TTHA1718 in sf9 cells are shown in Figures 1, 3A, and 4A, respectively.

For GB1 in sf9 cells, we could achieve unambiguous assignments for approximately 98% of the backbone 1H, 15N, 13C, and 13C resonances, as well as 76% of 1H, 42% of 1H, and 75% of 13C resonances of GB1 in sf9 cells, by analyzing 3D triple-resonance NMR spectra as well as 3D 15N- and 13C-separated NOESY spectra (Supporting Information, Figure S5). Moreover, approximately 67% of the side-chain 1H, 1H, and 1H resonance assignments were achieved by analyzing HCCH-TOCSY and spectra for an AILV selectively 13C/15N-labeled sample (for the experimental procedures see the Supporting Information).

3D structures were first calculated with the program CYANA using NOE information as well as backbone dihedral angle restraints derived from chemical shifts. The resulting structure was further refined with the assistance of Bayesian inference using the CYBAY module in CYANA,[6b,13] which was essential for the accurate structure determinations in sf9 cells. 189 NOE-derived distance restraints, including 54 long-range restraints, were used (Supporting Information, Table S1). The final 1900 conformers are well defined, with average backbone and side-chain root-mean-square deviations (RMSDs) of 0.51 and 0.85 Å relative to the mean coordinates (Figures 2A–C). The backbone and side-chain RMSDs between the mean structure in sf9 and that in diluted solution (RCSB Protein Data Bank (PDB) code: 2N9K) are...
1.61 and 2.42 Å, respectively. Figure S6 (Supporting Information) shows the corresponding RMSD per residue and its standard deviation over all the sampled conformers. RMSDs of the main-chain atoms were small for most residues, except for the residues 22–26 and 28 in the loop, and the α-helix that presented higher values around 1.5 Å (Supporting Information, Figure S6F, left panel). RMSDs of the side-chain atoms showed similar correlation with those of the backbone, and only the α-helix was slightly different to that in diluted solution (Supporting Information, Figure S6F, right panel). These residues coincide well with a region exhibiting chemical shift differences between the in-cell and diluted solution samples (residues 20–24, 27; Figure 2D). In sf9 cells, the relative position of the α-helix is tilted significantly away from the β-sheet (Figure 2E). This difference of the α-helix is yielded by altered NOE cross-peak patterns, particularly for NOEs observed between aliphatic and aromatic side-chains located between the α-helix and β-sheet (Supporting Information, Figures S6B–E).

The changes in chemical shift and 3D structure for this region are presumably because of the effects caused by the intracellular environment. It is likely that residues Ala21, Val22, Ala24, Ala25, and Ala27, which form a hydrophobic patch on the protein surface, interact with other molecules non-specifically, thus inducing the conformational difference. Previously we showed that the GB1 structure in E. coli cells has conformational differences in a similar region (residues 20–24) when compared with that in diluted solution. The conformational difference in this region has also been reported in a molecular dynamics study simulating crowded environments, in which the relative position of the α-helix was destabilized and opened to the solvent.

The structure determinations of Ub3A and TTHA1718 in sf9 cells were also performed from the NOE-derived distance restraints, for which the chemical shift assignments were transferred from the data in diluted solution based on the knowledge that chemical shift differences for these proteins between sf9 cells and diluted solution were small (Figures 3 and 4; Supporting Information, Figure S7). The resulting structure ensembles of Ub3A with 4400 conformers and TTHA1718 with 1000 conformers are well-defined with average backbone RMSD of 0.39 and 0.88 Å, respectively, with respect to the mean coordinates (Figures 3B and 4B). For Ub3A, the backbone RMSD between the mean structure and the structure in diluted solution is 1.31 Å. Figure S7B (Supporting Information) shows the RMSD of each residue compared to the structure in diluted solution, and its standard deviation over all sampled conformations. The regions comprising residues 17–21, 32–40, 46, and 52–60, as well as the flexible C-terminus, showed relatively large conformational deviations the structure in diluted solution. However, small chemical shift differences for these regions do not corroborate the conformational differences, suggesting that the difference for Ub3A may be because of limited distance restraints in these regions. For TTHA1718, the RMSDs for backbone and side-chain atom coordinates in sf9 cells, compared to the structures in diluted solution, were larger than in the cases of GB1 and Ub3A. This was particularly noticeable for the putative metal-binding loop region of residues 9–18 (Supporting Information, Figure S7D), for which NOE distance restraints were hardly collected, presumably because of an exchange process related...
to the binding of various metal ions. Excluding this region, the backbone RMSD to the structure in diluted solution drops to 1.27 Å.

CaM and HRas in sf9 cells exhibited heavily overlapped cross-peaks in 2D $^1$H-$^1$H NMR heteronuclear single-quantum correlation spectroscopy (HSQC; Supporting Information, Figures S8B–E). Therefore, we prepared samples with methyl- and aromatic-selective $^1$H/$^13$C-labeling, and assessed the feasibility of obtaining structural information based on 2D $^1$H-$^13$C HSQC and 3D $^13$C-separated NOESY spectra (Figure 5; Supporting Information, Figure S9). In both cases, well-resolved 3D $^13$C-separated NOESY spectra were acquired, indicating that our approach is effective for NOE-based structural analysis of proteins with molecular weight over 15 kDa in eukaryotic cells. Furthermore, comparison of in-cell NMR spectra of CaM with corresponding spectra in diluted solution (Figure 5A; Supporting Information, Figure S10) suggest that CaM in sf9 cells exists in a state similar to Mg$^{2+}$-bound CaM in diluted solution. This data indicates that the bioreactor system successfully suppressed the stress-induced Ca$^{2+}$ release, which has been reported in our previous work.

In-cell NMR spectra suggest that HRas in sf9 cells is in the "inactive" guanosine diphosphate (GDP)-bound state (Figure 5B; Supporting Information, Figure S11), which is reasonable considering that the C-terminal truncation prohibits the GDP to guanosine triphosphate (GTP) exchange at the cell membrane, while HRas-bound GTP will be hydrolyzed by its intrinsic GTPase activity.

In this article we confirmed that 3D structures of proteins of less than 10 kDa size can be determined exclusively from NOE-derived distance restraints acquired in living sf9 cells. Bayesian inference was effective for determining 3D structures with sufficient precision to detect conformational differences between the structures in sf9 cells and in diluted solution. Amongst three successful examples, we demonstrated that a GB1 structure in sf9 cells was determined with resonance assignments exclusively from in-cell NMR data. Our results extend the range of in-cell NMR spectroscopy, thus contributing to the understanding of the effects of intracellular molecular crowding to protein conformation and dynamics. A similar procedure can be utilized for proteins expressed in human cells.

For small proteins with a lower expression level in sf9 cells, it is expected that side-chain resonance assignments are not sufficient while well-resolved NOESY spectra can be obtained. This is because the rapid relaxation of transverse $^1$H and/or $^13$C magnetization reduces the sensitivity of triple-resonance NMR experiments significantly, whereas NOESY spectra are less affected. Methodological improvements in assigning backbone/side-chain resonances from NOESY spectra may broaden the range of applicable proteins in eukaryotic cells in the future. For medium-sized proteins, we confirmed that high-resolution structural information can be obtained from in-cell NOESY experiments in combination with selective $^1$H/$^13$C-labeling. Note that stable isotope-enrichment is the labeling method of proteins that least perturbs their physical properties when compared with chemical modifications or protein fusion. Our approach will therefore also be beneficial for the structural analysis of flexible regions on the protein surface or intrinsically disordered proteins in eukaryotic cells.

**Acknowledgements**

The authors thank Dr. Junpei Hamatsu for assistance with NMR measurements. We gratefully acknowledge financial support from the Funding Program for Core Research for Evolutional Science and Technology (CREST; JPJ1CR13M3) from the Japan Science and Technology Agency (JST), Next Generation World-Leading Researchers (NEXT Program), Grants-in-Aid for Scientific Research (JP16H04339 to M.M., JP15K06979 to T.I., JP17K07312 to P.G.), Challenging Exploratory Research (JP15K14494 to Y.I.), and Scientific Research on Innovative Areas (JP26102538, JP25120003, and JP16H00779 to T.I., JP15H01645, JP16H00847, and JP17H05887 to Y.I.) from the Japan Society for the Promotion of Science (JSPS).

**Conflict of interest**

The authors declare no conflict of interest.

**Keywords:** eukaryotic cells · macromolecular crowding · NMR structure determination · NOESY · proteins
How to cite: Angew. Chem. Int. Ed. 2019, 58, 7284–7288
Angew. Chem. 2019, 131, 7362–7366


Manuscript received: January 22, 2019
Revised manuscript received: March 20, 2019
Accepted manuscript online: April 1, 2019
Version of record online: April 25, 2019