Part 3
Basic In-cell NMR Techniques

CHAPTER 5

Protein Structure Determination in Living Cells from NOE-derived Distance Restraints

TEPPEI IKEYA,*[a] PETER GÜNTERT[a,b,c] AND YUTAKA ITO*[a]

[a] Department of Chemistry, Tokyo Metropolitan University, 1-1 Minami Osawa, Hachioji, Tokyo 192-0397, Japan; [b] Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany; [c] Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland
*Emails: tikeya@tmu.ac.jp; ito-yutaka@tmu.ac.jp

5.1 Introduction

In an intracellular environment, bio-macromolecules express their biological functions under extreme molecular crowding conditions[1] in which both specific and non-specific interactions with other proteins, nucleic acids, co-factors, ligands and various other small molecules play an important role. Although *in vitro* methods of structure determination of bio-macromolecules, X-ray crystallography, solution NMR spectroscopy, and single particle analysis with cryo-electron microscopy, have made significant contributions to understanding the structural bases of their biological activity, detailed *in vivo* observations of the behaviour of bio-macromolecules are still challenging. NMR spectroscopy targeting live cell samples, in-cell NMR,[2–5] is currently the only method for observing proteins and nucleic acids at atomic resolution in living biological systems.

New Developments in NMR No. 21
In-cell NMR Spectroscopy: From Molecular Sciences to Cell Biology
Edited by Yutaka Ito, Volker Dötsch and Masahiro Shirakawa

Published by the Royal Society of Chemistry, www.rsc.org

From the very initial stage of in-cell NMR studies, Dötsch and co-workers have proposed that the structural changes occurring in bio-macromolecules under intracellular crowding conditions are one of the important targets in-cell NMR spectroscopy has to tackle.[6] However, numerous problems arising from the properties of live cell samples, such as the short lifetime of cells in NMR tubes, severe signal overlap caused by much wider line-widths, background noise signals, sample inhomogeneity, *etc.* prevented the efficient determination of 3D structures of bio-macromolecules in living cells.

The first *de novo* protein 3D structure determined in living cells was accomplished for the *T. thermophilus* HB8 TTHA1718 gene product (henceforth referred to as TTHA1718), a putative heavy metal binding protein consisting of 66 amino acids, which was overexpressed in *E. coli* cells.[7] At that time, it was not clear whether the established solution NMR approaches for purified proteins,[8] in which 3D structures are calculated from nuclear Overhauser effect (NOE)-derived distance restraints based on backbone and side-chain resonance assignments, would be effective for proteins in living *E. coli* cells. Eventually, Sakakibara *et al.* demonstrated that the NOE-based approach is feasible for proteins in living cells.[7] In this study the motivation of Sakakibara *et al.* was to determine the 3D protein structure on the basis of chemical shifts and distance/angle information obtained exclusively in living cells, thus establishing a robust procedure applicable to various proteins. Generally speaking, triple-resonance NMR experiments, which are measured for resonance assignments, exhibited disappointing results due to the rapid relaxation of transverse ^{1}H, ^{13}C and ^{15}N magnetization caused by the increased rotational correlation time of proteins in cells. However, the backbone/side-chain assignment stage becomes very important for structural studies in cells, in particular for proteins experiencing large structural changes, for which the transfer of *in vitro* assignments to in-cell spectra becomes problematic. Sakakibara *et al.* demonstrated that their newly developed procedure was effective for TTHA1718 in *E. coli* cells by determining a high-resolution 3D structure ensemble with a backbone RMSD below 1.0 Å.[7] However, when applying a similar procedure to proteins with much lower intracellular concentrations in *E. coli* or in eukaryotic cells, further development and optimization of the methodology was necessary.[9]

Until now, NOE-based 3D structure determination of proteins in intracellular environments has been performed in *E. coli*[7,10] and in Sf9 cultured insect cells,[11] a clonal isolate of *Spodoptera frugiperda* Sf21 cells. Figure 5.1 summarizes the experimental NMR procedures used for TTHA1718 in *E. coli*[7] and in the study of *Streptococcus* protein G B1 domain (henceforth referred to as GB1) in *E. coli* cells.[10] The approach used for the case of GB1 in Sf9 cells[11] was essentially identical to that used for GB1 in *E. coli*. The procedure comprises six stages: (1) rapid acquisition of multidimensional (*n*D) NMR data, (2) reconstruction of NMR spectra, (3) manual assignment of backbone

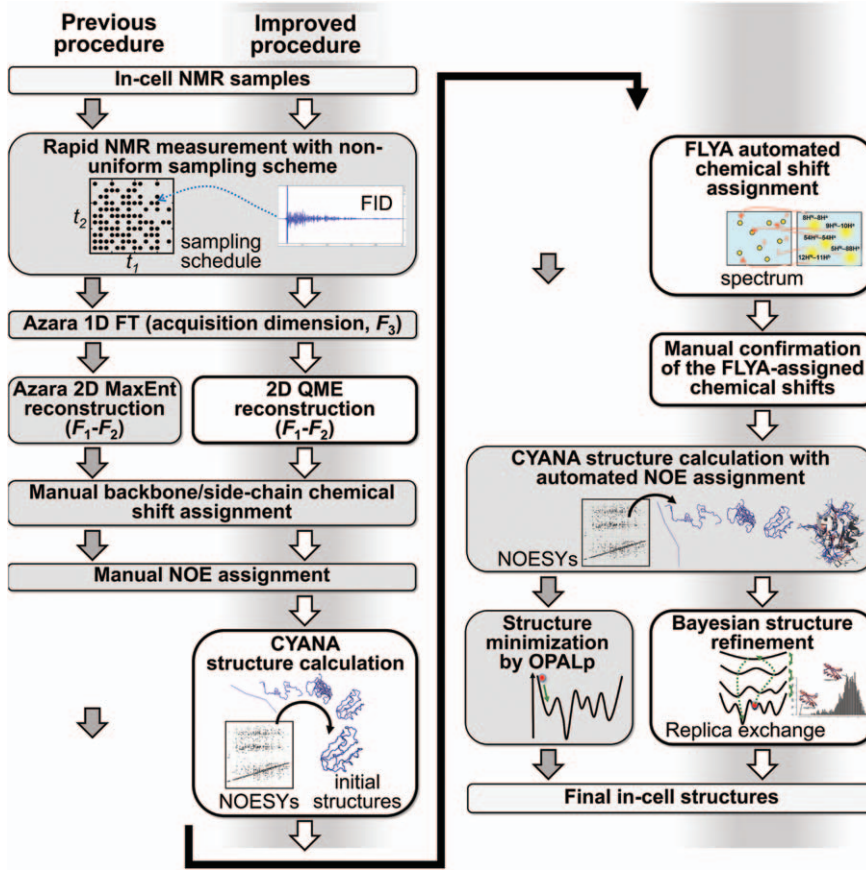


Figure 5.1 Procedures for in-cell NMR protein structure determination from NOE-derived distance restraints. The procedure used for the case of TTHA1718 in *E. coli* cells (left) and that used for the case of GB1 in *E. coli* or in Sf9 cells (right) are compared side-by-side. New and improved steps are highlighted: quantitative maximum entropy (QME) processing of non-uniformly acquired data, FLYA automated chemical shift assignment, and structure calculation with Bayesian inference-based refinement to obtain the final in-cell structures.
Adapted from ref. 10 [https://doi.org/10.1038/srep38312] under the terms of a CC BY 4.0 licence [https://creativecommons.org/licenses/by/4.0/].

and side-chain resonances, (4) manual assignment of NOEs, (5) structure calculations with automated NOE-assignment, and (6) structure refinement. Automated chemical shift assignment was optionally used during the structure calculation stage.

In this chapter we introduce the experimental approaches and discuss their use for probing protein 3D structure in living cells, and we also give an outlook to future areas of research and applications of cellular solution state NMR spectroscopy.

5.2 Experimental Procedures

5.2.1 Host Cells

The experimental procedures for in-cell NMR studies using *E. coli* and Sf9 cells are described in Chapters 2 and 4, respectively. The expression level of the TTHA1718 and GB1 proteins in *E. coli* cells resulted in effective concentrations of 3–4 mM[7] and 250 μM[10] in NMR tubes, respectively. In Sf9 cells, GB1, TTHA1718 and human ubiquitin with the three alanine mutations L8A, I44A, and V70A (henceforth referred to as Ub3A), for which 3D structures were determined from NOE-derived distance restraints, were expressed with concentrations of 100–200 μM in NMR tubes.[11] These data suggest that, currently, an effective concentration of ~100 μM is the lower limit for in-cell protein structure determination from NOEs. Similar NOE-based structure determination in *Xenopus laevis* oocytes should also be possible, since it has been reported that proteins can be introduced at up to ~0.7 mM intracellular concentration.[12]

Requiring a cell viability of 85% or more, the time limits for continuous in-cell NMR measurements are ~6 hours for *E. coli*[7] or ~3 hours for Sf9 cells.[13] Introduction of a bioreactor system that supplies fresh medium into the NMR tube continuously[14,15] (see Chapter 7) has been shown to be very effective for extending the time limit significantly, enabling the recording of highly sensitive in-cell NMR experiments.[11]

5.2.2 Labelling Strategy

With both *E. coli* and Sf9 cells, uniformly $^{13}C/^{15}N$-labelled samples were prepared for 3D triple-resonance NMR experiments and 3D NOESYs.[7,10,11,16] In contrast to 3D ^{15}N-separated NOESY, 3D ^{13}C-separated NOESY spectra measured on uniformly ^{13}C-labelled samples were difficult to analyse, because a considerable number of 'background' cross-peaks were present (Figure 5.2a). In order to solve this problem, selective protonation of methyl groups[17] was performed against an ^{2}H-background. For structure determination of TTHA1718 in *E. coli* cells, Ala/Leu/Val-selectively methyl $^{1}H/^{13}C$-labelled samples (TTHA1719 has no Ile residues) were prepared by incubating with [3-^{13}C] alanine and [U-^{13}C, 3-^{2}H] α-ketoisovaleric acid, a precursor of leucine and valine, for acquiring 2D $^{1}H–^{13}C$ HSQC and 3D $^{13}C/^{13}C$-separated HMQC-NOE-HMQC spectra.[7,16] Two additional methyl-$^{1}H/^{13}C$-labelled samples (Ala/Val- and Leu/Val-selectively $^{1}H/^{13}C$-labelled) were also utilized for amino acid classification of methyl $^{1}H–^{13}C$ correlation cross peaks. The selective methyl $^{1}H/^{13}C$-labellings turned out to be very effective for obtaining well resolved 3D NOESY spectra.

The quality of 3D $^{13}C/^{13}C$-separated HMQC-NOE-HMQC spectra can be further improved by suppressing strong intraresidual methyl-methyl NOEs, which sometimes prevented the collection of weak long-range NOEs. In the case of GB1 in *E. coli*, Ala/Ile/Leu/Val-selectively methyl $^{1}H/^{13}C$-labelled

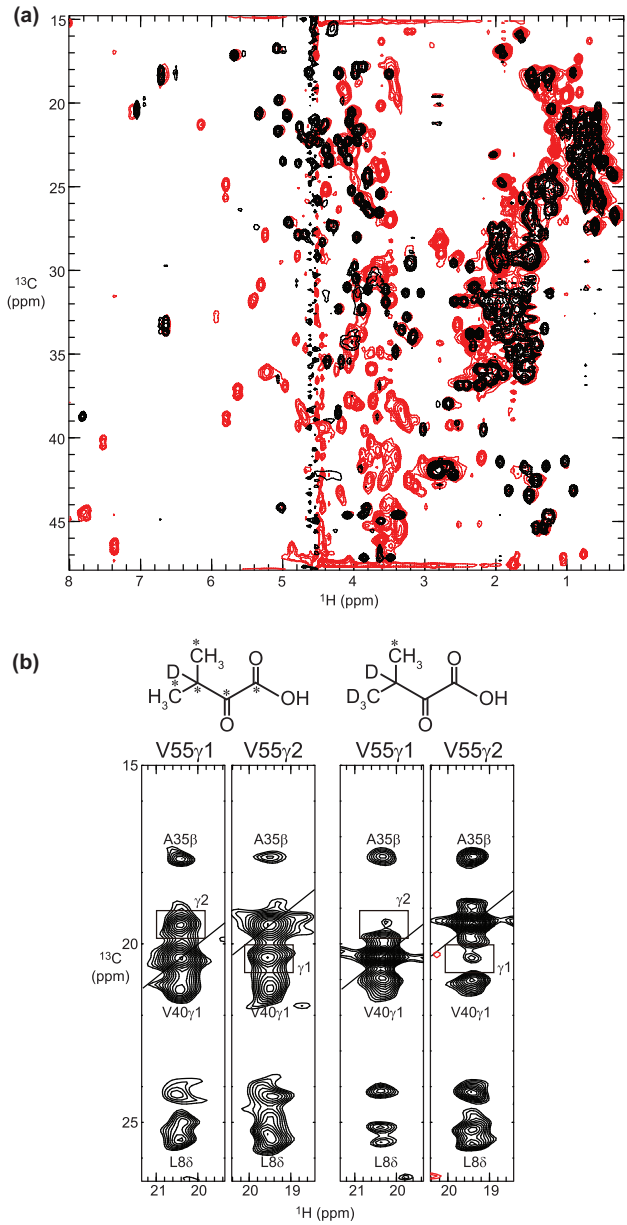


Figure 5.2 (a) Background ^{1}H–^{13}C correlation cross peaks originating from uniform ^{13}C-labelling. Overlay of the ^{1}H–^{13}C HSQC spectra of purified TTHA1718 (black) and *E. coli* cells expressing TTHA1718 (red). (b) ^{13}C–^{13}C cross-sections corresponding to the ^{1}H frequencies of the Val-55 γ1 and γ2 methyl groups of GB1 extracted from the 3D ^{13}C/^{13}C-separated HMQC-NOE-HMQC spectra measured on Ala/Ile/Leu/Val-selectively methyl ^{1}H/^{13}C-labelled GB1 in *E. coli* cells prepared using [U-^{13}C, 3-^{2}H] α-ketoisovaleric acid (left) or [3-methyl-^{13}C, 3,4,4,4-^{2}H4] α-ketoisovaleric acid (right). Intraresidual NOEs are indicated by boxes.
(a) Reproduced from ref. 7 with permission from Springer Nature, Copyright 2009.

samples were prepared using [3-^{13}C] alanine, [methyl-^{13}C, 3,3-^{2}H2] α-ketobutyric acid, a precursor of isoleucine, and [3-methyl-^{13}C, 3,4,4,4-^{2}H4] α-ketoisovaleric acid, which yielded 3D ^{13}C/^{13}C-separated HMQC-NOE-HMQC spectra with improved quality due to the absence of intraresidual methyl-methyl NOEs and $^{1}J_{CC}$ scalar couplings in ^{13}C-dimensions (Figure 5.2b). For the 3D structure determination of GB1 in Sf9 cells, selective ^{1}H/^{13}C-labelling was also applied for aromatic residues: a total of 10 differently labelled samples (Ala/Leu/Phe/Thr/Trp/Tyr/Val- (ALVTFWY-), Ala/Ile/Leu/Val- (AILV-), Ala-, Ile-, Leu-, Phe-, Thr-, Trp-, Tyr-, and Val-selective labelling were prepared) utilizing a medium based on the composition of IPL-41 (Thermo Fisher Scientific).[11]

5.2.3 Rapid *n*D NMR Measurement and the Reconstruction of Spectra

As was mentioned above, in-cell NMR studies inherently suffer from short time limits for experiments, thus requiring a large reduction in measurement times from the 1–2 days conventionally spent for each 3D experiment. Moreover, multiple scans are required from the signal-to-noise ratio (S/N) point of view, since in-cell NMR studies always face a problem of low sensitivity due to the lower intracellular concentration of biomacromolecules of interest. Recently, various new acquisition schemes for *n*D NMR experiments have been proposed for dramatic improvements in both sensitivity and resolution.[18,19] One of these approaches, projection reconstruction,[20] has been applied successfully to the backbone resonance assignment of GB1 in *E. coli*.[21]

In the 3D protein structure determination studies in *E. coli* and in Sf9, non-uniform sampling (NUS, also known as nonlinear sampling) methods[22–24] were employed. NUS has been shown to be a robust technique, in which sampling points on the Nyquist grid for indirect dimensions are decimated (Figure 5.3a). Since discrete Fourier transform (DFT) cannot be used for processing NUS-sampled data, various 'non-FT' methods have been developed for the reconstruction of *n*D NMR spectra. When NUS was applied for the first time to 2D NMR experiments, an exponential weight function was employed for selecting sampling points in the decaying indirect dimension.[22] Later, Wagner, Hyberts and co-workers contributed significantly to the application of NUS to 3D and 4D NMR experiments.[23]

Rovnyak *et al.* reported that optimal resolution for indirect dimensions can be achieved with an acquisition time of 3×transverse relaxation time (T_2) of the observed nuclei, while optimal S/N can be achieved with an acquisition time of 1.6×T_2.[25] In contrast to the conventional acquisition scheme, in which data sampling is rarely performed beyond 0.4×T_2, the NUS scheme can extend acquisition times for indirect dimensions without excessively long measurement time.

For in-cell NMR studies, optimal resolution in ^{15}N-dimensions (with 2000 Hz spectral width) would be achieved with 120–150 ms acquisition time, corresponding to 240–300 complex points, since ^{15}N T_2 relaxation times for TTHA1718 in *E. coli* cells have been reported to be in the range of 40–50 ms.[7] Similarly, 64–80 ms acquisition time is required for optimal S/N for ^{15}N dimensions, corresponding to 128–160 complex points. Note that these optimal acquisition times are significantly shorter than commonly used for *in vitro* experiments.

For TTHA1718 in *E. coli* cells, a fresh sample was prepared for each experiment and an NUS scheme was applied for the indirectly acquired dimensions in combination with 2D maximum entropy (MaxEnt) processing using the Azara package[26] to provide considerable time savings.[7] In order to achieve NUS sampling, the pulse sequences were modified according to the procedure reported by Rovnyak *et al.*[24] With this technique the duration of each 3D experiment was reduced to 2–3 h. To ensure that only data from 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 sets were combined to generate a new data set with improved S/N ratio until the 2D ^{1}H–^{15}N HSQC spectra showed marked changes (Figure 5.3b).

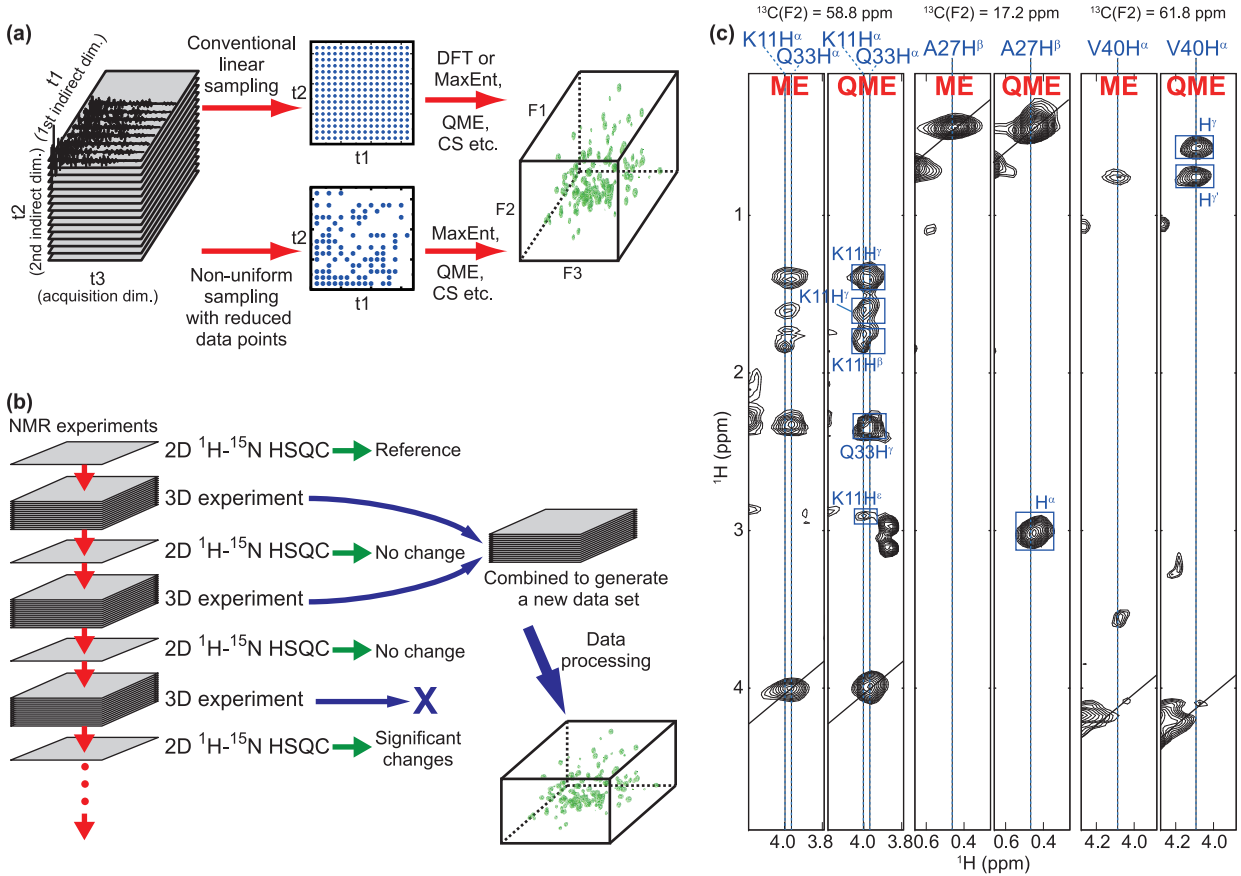
For TTHA1718 in *E. coli* cells, it was found that 2D MaxEnt processing with Azara sometimes provided unsatisfactory spectra with missing cross peaks, particularly for the experiments including strong and sometimes very sharp background signals. For GB1 in *E. coli* cells, where a similar or even more severe problem was expected due to the much lower intracellular concentration of GB1 compared to TTHA1718, a novel quantitative maximum entropy (QME) reconstruction method was employed.[13]

The QME software differs from existing MaxEnt routines, such as that in the Azara package, by its choice of feasibility criteria, the entropy function that is maximized, and the iterative search method employed. QME implements a holistic reconstruction scheme, by which the entire spectrum is reconstructed in one iteration, rather than a series of consecutive steps (*i.e.* 1D slices for a 2D spectrum, or 2D planes for a 3D spectrum). For most of the *n*D NMR spectra obtained *in vitro*, MaxEnt and QME provided very similar results. However, QME improved significantly the quality of in-cell NMR spectra, which suffer from strong background signals from endogenously expressed proteins and a wide dynamic range in peak intensity (Figure 5.3c).[10,13] NUS-sampled data can also be processed with other algorithms, such as L_p-norm ($0 < p \leq 1$) minimization, referred to as compressed sensing, which was recently introduced to the NMR field.[27,28]

5.2.4 Backbone and Side-chain Resonance Assignment

Backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C'$, and side-chain $^{13}C^{\beta}$ resonance assignments of TTHA1718 and GB1 in living *E. coli* cells were performed by analysing six 3D triple-resonance NMR spectra: HNCA, HN(CO)CA, CBCA(CO)NH, CBCANH,

Published on 09 December 2019 on https://pubs.rsc.org | doi:10.1039/9781788013079-00063



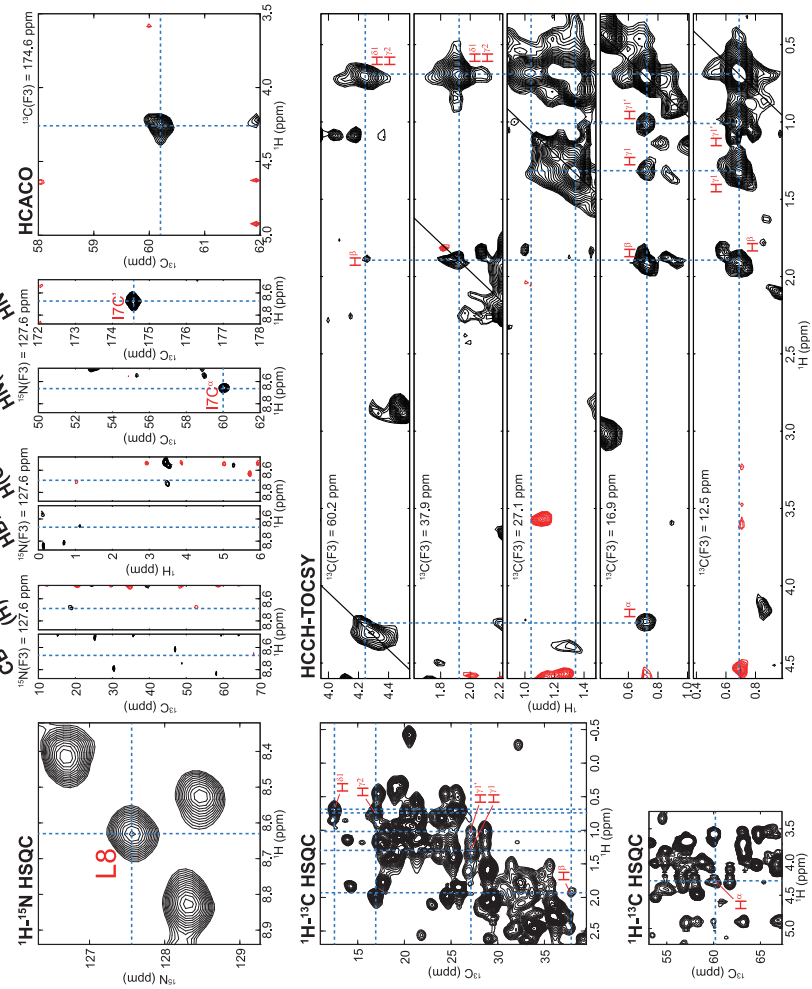
HNCO, and HN(CA)CO. Virtually complete backbone assignments were achieved for both TTHA1718 and GB1 in *E. coli* cells.[7,10]

A similar strategy was employed for GB1 in Sf9 cells. In the first report, Hamatsu *et al.* succeeded in assigning approximately 80% of the backbone NMR resonances of GB1.[13] The assignment was incomplete due to the lower intracellular concentration of GB1 and shorter lifetime of Sf9 cells in NMR tubes when compared to *E. coli* cells. In the following study, Tanaka *et al.* introduced a bioreactor system and succeeded in prolonging the lifetime of Sf9 cells in NMR tubes to be comparable to that under ''optimal'' conditions in culture flasks, maintaining >90% cell viability as well as protein stability in the cells for at least 24 hours. Consequently, unambiguous assignments could be achieved for approximately 98% of the backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, and $^{13}C'$ resonances.[11]

Backbone $^{1}H^{\alpha}$ and side-chain ^{1}H and ^{13}C resonances of TTHA1718 in *E. coli* cells were assigned by analysing three 3D triple-resonance NMR spectra: HBHA(CBCACO)NH, H(CCCO)NH, and (H)CC(CO)NH. The chemical shifts of the majority of H^{α} and H^{β}, and ~1/3 of the other aliphatic $^{1}H/^{13}C$ side-chain resonances were determined.[7] GB1 in *E. coli*, in contrast, suffered from lower sensitivity in these 3D triple-resonance experiments, which lacked many of the expected cross peaks. Ikeya *et al.* therefore measured a 3D HCACO spectrum for the additional assignment of $^{1}H^{\alpha}$ resonances, and 3D HCCH-COSY and HCCH-TOCSY spectra for the side-chain resonance assignment (Figure 5.4). In combination with the amino acid-type classification of methyl $^{1}H–^{13}C$ correlation cross peaks from the in-cell NMR spectra of methyl-selectively $^{1}H/^{13}C$ labelled samples, the chemical shifts of 88% of $^{1}H^{\alpha}$, 71% of $^{1}H^{\beta}$, and 32% of the other aliphatic $^{1}H/^{13}C$ side-chain resonances (including ~88% of the methyl resonances) of GB1 in *E. coli* cells were assigned by manual analysis.[10]

A similar side-chain assignment strategy was employed for GB1 in Sf9 cells, for which CBCA(CO)NH and HBHA(CBCACO)NH spectra of uniformly $^{13}C/^{15}N$-labelled samples and HCCH-TOCSY spectra of AILV-selectively $^{13}C/^{15}N$-labelled sample were analysed.[11]

Figure 5.3 (a) Rapid acquisition of 3D NMR spectra using a non-uniform sampling scheme. (b) Repeated observation of 3D NMR spectra with intermittent monitoring of the sample condition by short 2D $^{1}H–^{15}N$ HSQC experiments. (c) Comparison of 3D NMR spectra of GB1 in *E. coli* cells processed with QME or MaxEnt reconstruction. $F_1(^{1}H)$-$F_3(^{1}H)$ slices of 3D HCCH-TOCSY at ^{13}C frequencies of 58.8, 17.2, and 61.8 ppm are shown from 2D MaxEnt and 2D QME reconstructed spectra for which the raw data were acquired using a non-uniform sampling scheme. Cross peaks are clearly visible at the ^{13}C frequency of 58.8 ppm, but largely absent for the other two ^{13}C frequencies.
(a) and (b) Reproduced from ref. 7 with permission from Springer Nature, Copyright 2009. (c) Reproduced from ref. 10 [https://doi.org/10.1038/srep38312] under the terms of a CC BY 4.0 licence [https://creativecommons.org/licenses/by/4.0/].



5.2.5 3D NOESYs

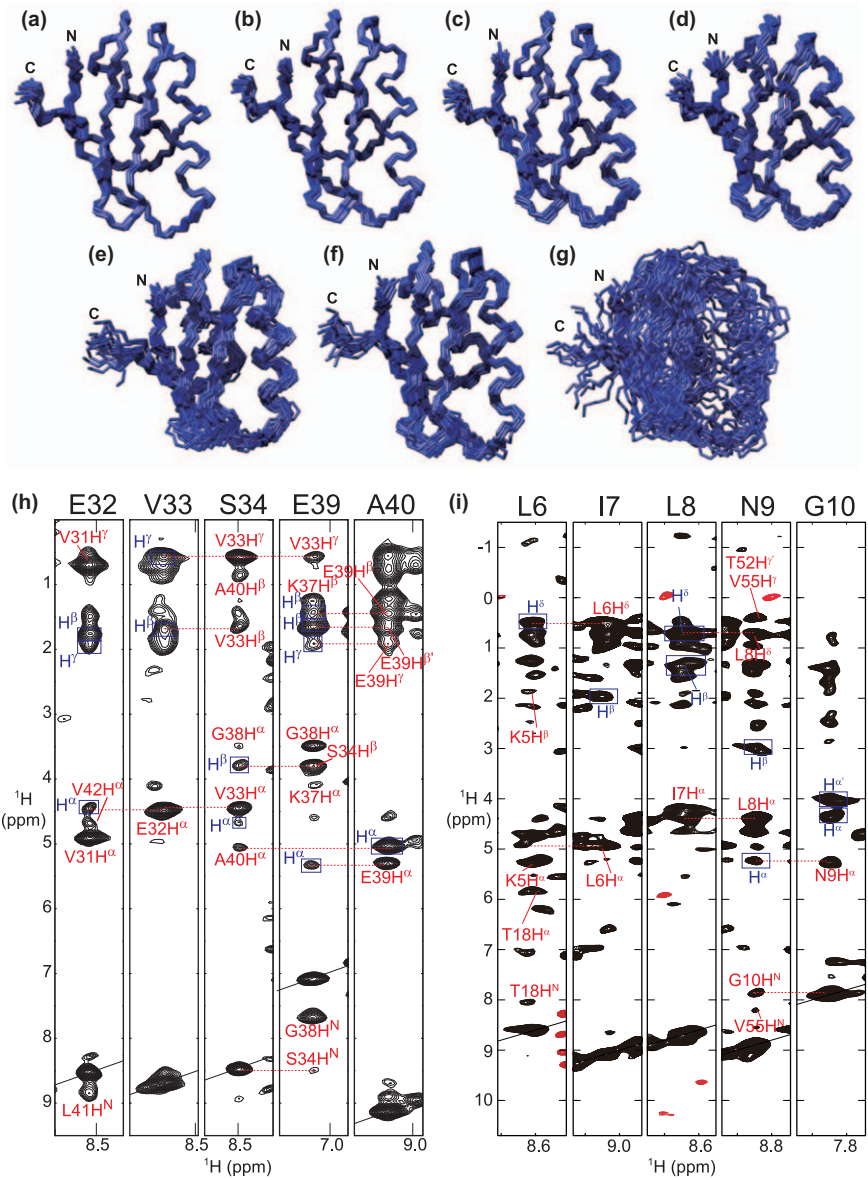
NUS was also employed for 3D NOESY spectra, with selecting typically 1/4–1/8 of the data points in the indirect dimensions randomly. Employing the non-FT methods to the NOESY-type experiments remains controversial because these spectra require accurate signal intensities for obtaining interatomic distances, which may not be ensured with NUS, especially in the case of signals with a high dynamic range.

Thus, as a preparatory analysis for determining the 3D structure of TTHA1718 in *E. coli* cells by the NOE-based approach, the effect of the artefacts arising from employing NUS and 2D MaxEnt processing to 3D NOESY-type spectra was evaluated by calculating structures from distance restraints obtained in simulated 3D NOESY spectra with various non-uniformly selected sampling points (Figure 5.5a–g).[29]

Artefacts expected from NUS and 2D MaxEnt processing were (1) distorted peak intensities, (2) the loss of original cross peaks, and (3) the emergence of false cross peaks. The second problem is, to a certain extent, inevitable when utilizing NUS, since the reduction of the total acquisition time causes a decrease in the S/N ratio of the data. The third problem is not likely to have happened since the 2D MaxEnt iteration process with the Azara software starts from a ''flat'' spectrum. Even if it happened, the false peak problem may not be significant because the procedure of resonance assignment compares several spectra simultaneously, and one can easily eliminate artificial cross peaks that are inconsistent with those in other spectra. The automated NOE assignment algorithm also systematically excludes ''orphan'' cross peaks.[30] The first problem, deviation of intensities for NOE cross peaks from ''real'' values is usually not so harmful to structure calculations, since the NOE intensities are generally interpreted as distance restraints with relatively large tolerances and the $1/r^6$-relationship between NOE intensities and distances reduces the relative error of the latter six-fold. Nevertheless, the miscalibration of intensities eventually became significant in the case of strongly reduced numbers of data points. There was a tendency that the more data points were omitted, the more the intensities of weaker cross peaks were underestimated. This resulted in a drop in the numbers of picked NOE cross peaks in the

Figure 5.4 Side-chain resonance assignment of GB1 in *E. coli* cells using exclusively 2D/3D in-cell NMR spectra. The manual assignment process for side-chain resonances of Ile-7 is illustrated. Whilst CBCA(CO)NH, (H)CC(CO)NH, HBHA(CBCACO)NH and H(CCCO)NH spectra corresponding to the ^{15}N frequency of Leu-8 (127.6 ppm) showed no cross-peaks, $^{13}C^{\alpha}$ and $^{13}C'$ resonances were observable in the HN(CO)CA and HNCO spectra, respectively, which were used for the assignment of the $^{1}H^{\alpha}$ resonance of Ile 7 in the $F_1(^{13}C)$-$F_3(^{1}H)$ slice of the HCACO spectrum. The analysis of the HCCH-TOCSY spectrum based on the $^{1}H^{\alpha}$ and $^{13}C^{\alpha}$ resonance assignments identified all side-chain $^{1}H/^{13}C$ resonances of Ile-7. The assignments are shown in the 2D ^{1}H–^{13}C HSQC spectrum.
Reproduced from ref. 10 [https://doi.org/10.1038/srep38312] under the terms of a CC BY 4.0 licence [https://creativecommons.org/licenses/by/4.0/].

spectra, and consequently poorer convergence of the structure calculations. Nevertheless, when sampling the same number of FIDs, the convergence was much worse when the data were acquired uniformly. This suggests that NUS is advantageous for acquiring 3D NOESYs, as compared to the conventional uniform sampling scheme, particularly for short samples with a short lifetime or in the case of low sensitivity. In addition, the results showed that poor convergence in the structure calculations can be improved by supplementing



some ''critical'' selectively measured long-range distance restraints, *e.g.* NOE-derived distance restraints involving side-chain methyl groups.

Figure 5.5h and 5i shows ^{1}H–^{1}H cross-sections extracted from the 3D ^{15}N-separated NOESY spectra of TTHA1718 in *E. coli* cells (b) and GB1 in Sf9 cells (c). With the help of NUS and MaxEnt or QME processing, high-quality 3D NOESY spectra were obtained within the lifetime of cells.

5.2.6 FLYA-automated Assignment

For GB1 in *E. coli* cells, the ^{1}H/^{13}C chemical shifts of side-chain resonances were further assigned based on NOESY data using the FLYA algorithm for automated resonance assignment.[31] It had been shown earlier[32,33] that FLYA can also be used exclusively with NOESY peak lists as input. This approach enabled side-chain assignments even where fast transverse relaxation made it impossible to collect enough signals from through-bond spectra such as, for instance, H(CCCO)NH, whereas NOESY spectra provided a significant number of side-chains signals. While it is difficult to unambiguously assign

Figure 5.5 TTHA1718 structures calculated from simulated 3D NOESY data with various nonlinear sampling points (a–g). The simulated raw data were obtained by omitting data points from the 3D ^{13}C-separated and ^{15}N-separated NOESY-HSQC spectra of TTHA1718 *in vitro*. The original 3D raw data have 512, 128 and 32 complex points for t_3 (^{1}H), t_1 (^{1}H) and t_2 (^{13}C or ^{15}N) dimensions, respectively. The analysis of NOE cross peaks in these simulated spectra and structure calculations were performed by employing essentially identical protocols as for the structure determination of TTHA1718 in *E. coli* cells.[7] In addition to the NOE-derived distance restraints, backbone torsion angle restraints and distance restraints for hydrogen bonds, which were used for the case of TTHA1718 in *E. coli* cells, were introduced. This figure shows the superimposed 20 final structures of CYANA calculations from six sets of complete or sparse 3D ^{13}C-separated and ^{15}N-separated NOESY spectra: (a) 64 and 32 uniformly sampled complex points for t_1 (^{1}H) and t_2 (^{13}C or ^{15}N), respectively. (b) 1024 randomly selected complex points out of the sampling space with 64 (t_1, ^{1}H)×32 (t_2, ^{13}C or ^{15}N) complex points. (c) 512 randomly selected complex points. (d) 256 randomly selected complex points. (e) and (f) 128 randomly selected complex points. (g) 16 and 8 uniformly sampled complex points for t_1 (^{1}H) and t_2 (^{13}C or ^{15}N), respectively. For the structure calculations presented in (f), NOE-derived distance restraints involving side-chain methyl groups, which were used for the case of TTHA1718 in-cell, were added. ^{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 spectra of TTHA1718 in *E. coli* cells (h, processed with MaxEnt) and GB1 in Sf9 cells (i, processed with QME). The cross-peaks due to interresidual NOEs are assigned in red. Intraresidual NOEs are indicated by blue boxes and annotated, H. (h) Reproduced from ref. 7 with permission from Springer Nature, Copyright 2009. (i) Adapted from ref. 11 with permission from John Wiley and Sons, Copyright © 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

resonances from the large number of candidates in NOESY spectra by conventional manual spectrum analysis, the automated FLYA approach can efficiently analyse all NOESY correlations in order to validate objectively proposed resonance assignments from the manual approach.

The FLYA algorithm[31] first generates expected peak lists on the basis of the amino acid sequence and the magnetization transfer pathways for the given spectra, which are stored in a general library. In the absence of a 3D structure, expected peaks for NOESY experiments are only predicted for pairs of atoms that are close in sequence. Optionally, expected long-range NOEs can be generated in addition if a 3D structure of the protein is provided. Assuming that the probability of a peak to be actually measured in the experiment decreases with the distance between the respective atoms, distance-dependent observation probabilities are assigned to the expected NOESY peaks. FLYA then determines the resonance assignment of the protein by optimally mapping expected peaks to experimentally measured peaks using a combination of an evolutionary optimization algorithm and a local optimization routine. Each expected peak can be mapped to only one measured peak. The evolutionary algorithm works with a population of individuals that represent an assignment for the entire protein. Assignments are optimized such as to be consistent with general chemical shift statistics from the Biological Magnetic Resonance Data Bank (BMRB) and such that corresponding chemical shift coordinates of different measured peaks assigned to the same atom agree within a given tolerance. The local optimization routine is applied to the individuals of each generation and in order to reassign the expected peaks for single atoms, which is repeated for a specified number, by default 15 000, of iterations. The locally optimized solutions are then recombined to form a new generation using a scoring function that takes into account the chemical shift statistics, the alignment of peaks assigned to the same atom, the completeness of the assignment, and a penalty for chemical shift degeneracy. The solution with maximal score value at the end of the calculation represents the final assignment of the protein. A number (typically 20) of independent runs of the evolutionary algorithm are performed with identical input data but different random number generator seeds, resulting in 20 chemical shift values for each atom. From these, a consensus chemical shift value and a measure of the self-consistency of the assignment are computed. This allows to increase the accuracy of the assignment, and to obtain a reliability measure for each assigned atom. Assignments with high self-consistency (''strong'' assignments) have been shown to be much more reliable than other (''weak'') assignments.[30]

The automated FLYA approach based on NOESY spectra as well as the spectra used for the manual side-chain resonance assignment, yielded assignments for an additional 48 ^{1}H, 1 ^{15}N, and 52 ^{13}C resonances for GB1 in *E. coli*,[10] which, in turn, provided additional distance restraints resulting in clearly improved in-cell structures that are closer to the *in vitro* one. Automated resonance assignment with FLYA is thus expected to be generally

advantageous, especially when only an insufficient number of signals can be observed in conventional through-bond spectra for side-chain resonance assignment.

5.3 Structure Determination

5.3.1 Conventional Approach for Protein Structure Determination

For TTHA1718 in *E. coli* cells, a sufficient number of NOE distance restraints could be collected from 3D ^{15}N- and ^{13}C-separated NOESY spectra measured on uniformly ^{13}C/^{15}N-labelled samples and 3D ^{13}C/^{13}C-separated HMQC-NOE-HMQC spectra measured on methyl-selectively ^{1}H/^{13}C-labelled samples. Backbone ϕ/ψ torsion angle restraints were obtained using the program TALOS.[34] At that time, indirect distance information was introduced as hydrogen bond restraints in the β-sheet region where the existence of hydrogen bonds was strongly suggested by interstrand NOEs. On the basis of these restraints, the 3D structure determination was performed with the program CYANA, which is widely used for NMR structure calculation and automatic chemical shift assignment as well as occasionally for molecular modelling in the *de novo* design of proteins.[35] The standard procedure of structure determination by CYANA consists of a conformational search with simulated annealing (SA) by molecular dynamics simulation (MD), which aims at satisfying ranges of distances and dihedral angles derived from experimental data, and subsequent structure optimization in a physical force field. CYANA uses torsion angle molecular dynamics (TAMD) that enables longer step-sizes than MD in Cartesian coordinate space and thus yields converged structures more quickly.[36] For in-cell structure determination, initially standard structure calculations with CYANA were used. Starting from 100 conformers with random torsion angles, seven cycles consisting of automated NOE assignment and structure calculation were run so as to gradually assign more NOE distance restraints and to converge the structures. In order to calibrate interproton distances from NOE peak intensities, many methods of NMR structure determination employ a simple approximation that categorizes peak intensities into a few classes with certain distance ranges, *e.g.* short-, middle-, and long-distance restraints. As another simple approach to obtain NOE distances, CYANA determines the calibration constant for a $1/r^6$ intensity-to-distance relationship by assuming that the median cross peak intensity in a NOESY data set corresponds to a given distance, with a default value of 4.0 Å, and was confined to the range 2.4–5.5 Å. Even using these simple approximations, the collection of a sufficient number of distance restraints compensates for the ambiguity of individual data, and this standard method of NMR structure determination was able to determine adequately accurate structures. Thus, the CYANA structure calculation yielded a well-converged structure ensemble of TTHA1718 in *E. coli* cells with a backbone RMSD of 0.96 Å to the mean

coordinates that were also in close agreement with the structure that had been determined independently *in vitro* from a purified sample.[7]

5.3.2 Bayesian Inference-assisted Structure Refinement

Later, as already mentioned, the procedure for *de novo* in-cell protein structure determination was improved by Bayesian inference-assisted structure refinement,[10] which yielded accurate 3D protein structures even from the severely limited number of distance restraints that can be obtained with much lower intracellular protein concentrations. Indeed, this method along with QME processing and FLYA automated assignment enabled the structure determination of the protein GB1 in living *E. coli* cells at an order of magnitude lower concentration (approximately 250 μM) in the NMR tubes than in the original report for TTHA1718 (3–4 mM). The NMR structure determination based on a Bayesian framework, so-called inferential structure determination (ISD), was originally proposed for *in vitro* data[37,38] and then adopted for in-cell NMR.[10] While the conventional structure determination uses a target function composed of the sum of molecular potential energy, and residuals between predicted and observed data, the Bayesian approach employs the posterior probability based on Bayes' theorem as a scoring function. Unlike the conventional target function, which is used for finding an optimal solution, the probability allows to quantify uncertainty and incompleteness from experimental errors and unknown physical factors. Since Bayesian inference generally requires an extensive search in the space of explanatory variables, ISD employs replica-exchange Monte Carlo (REXMC)[39,40] for efficient sampling. While the original ISD approach had achieved considerable success,[38,41] it was not sufficient for sparse data in-cell NMR due to the high background content and low S/N ratio in the spectra. For the efficient analysis of in-cell NMR data along with more ideal prior information, the CYBAY algorithm was developed within the framework of CYANA, containing automatic NOESY cross peak analysis, fast global conformational search by TAMD, and structure optimization by REXMC with the physical force field.[42] CYANA was also equipped with the Amber ff03 physical force field[43] and a generalized Born (GB) implicit water model[44] optimized for TAMD, which provides more detailed prior probability distributions by sampling conformations more accurately on the energy landscape of proteins in torsion angle space. TAMD in CYANA permits longer time steps than Cartesian space MD simulation, and thus achieves a faster and wider conformational search. Moreover, CYBAY handles ambiguous NOE assignments in the calculation. As a result, CYBAY permitted more accurate data-driven structure determination with the sparse data from in-cell NMR.

For the structure determination of GB1 in *E. coli* cells, 1900 conformers were obtained as the final structures from the calculation, which employed a sufficient number of 10^7 REXMC steps[10] featuring frequent exchanges among the 10 different runs (replicas) at all temperatures used. Figure 5.6a

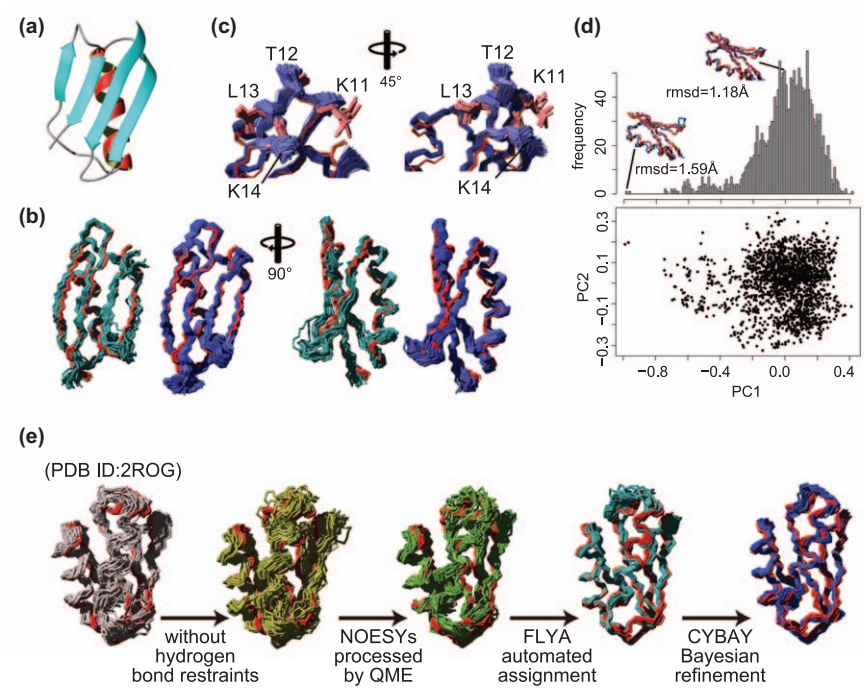


Figure 5.6 NMR structure of the protein GB1 in living *E. coli* cells. (a) Ribbon diagram of the structure with the highest posterior. (b) In-cell GB1 structures obtained by CYBAY (blue) and conventional CYANA calculation with the FLYA automatic chemical shift assignment (light blue), showing the backbone (N, C^{α}, C′) atoms. A total of 380 (20%) out of 1900 CYBAY conformers and 20 out of 100 in the final step of the conventional method are superimposed to the 20 structures determined *in vitro* (red), respectively. (c) Superpositions of the 20 GB1 structures determined *in vitro* (red) and the ensemble of in-cell CYBAY structures (blue), showing the side-chains of residues 11–14. (d) Distributions of the first principal component (top) and the first and second ones (bottom). (e) Improved structure calculations of the protein TTHA1718 in living *E. coli* cells. Previously reported structure of TTHA1718 in living *E. coli* cells[7] computed with hydrogen bond restraints (grey). Structure calculated without hydrogen bond restraints (yellow). Structure obtained with the NOESY spectra newly processed by QME (green). Structure obtained using QME-processed spectra and additionally automatically assigned chemical shifts by FLYA (cyan). Structure obtained using QME-processed spectra, FLYA automated assignments, and CYBAY Bayesian refinement (blue). For comparison, the structure determined *in vitro* is shown in red in all panels. All structures are represented by bundles of 20 CYANA conformers in (a–d) and 360 (20%) out of 1800 CYBAY conformers in (e), showing the backbone (N, C^{α}, C′) atoms.
Adapted from ref. 10 [https://doi.org/10.1038/srep38312] under the terms of a CC BY 4.0 licence [https://creativecommons.org/licenses/by/4.0/].

shows the representative CYBAY-refined structure with maximal *a posteriori* (MAP) estimation of GB1 in living cells. One of the advantages of Bayesian inference is that it provides not only the best structure with the lowest target function value (or MAP), but also distributions that reflect the uncertainty of experimental data such as measurement errors and shortage of information. Performing principal component analysis (PCA) of the coordinates of the ensemble structures, a slightly non-normal distribution was observed along the first principal component (PC1), which suggested the presence of a small number of minor populations in the vicinity of the major region (Figure 5.6d). Whereas the major region comprises structures within about 1.2 Å RMSD from the *in vitro* structure that had been determined independently by the conventional SA method, the minor populations included structures with approximately 1.6 Å RMSD (Figure 5.6d). Overall, the CYBAY structure ensemble with 1900 conformers was well defined with an average backbone RMSD of 0.43 Å to the mean coordinates (Figure 5.6b).

For the first in-cell structure determination of the protein TTHA1718, backbone hydrogen bond restraints were added for the β-sheet and α-helical regions where their existence was explicitly indicated by NOEs. General NMR structure determinations often employ this approach, but it may miss slight deviations from canonical secondary structure manifested in the experimental data. It was instructive to improve the TTHA1718 structures by the Bayesian approach of data-driven structure determination with prior information. Thus, CYBAY Bayesian-assisted structure optimization, as well as QME data processing and FLYA automatic resonance assignment, was applied to the previously recorded NMR data of TTHA1718 in living *E. coli* cells.[7] Omitting the hydrogen bond restraints that had been used in the first in-cell structure determination, the structure obtained with CYBAY was obviously different from the original (Figure 5.6e). The QME reconstruction clearly enhanced the intensities of numerous cross peaks of NOESY spectra, and revealed previously undetected signals. Besides, FLYA automatic chemical shift assignment achieved additional resonance assignments. Overall, 608 NOE-derived distance restraints, including 188 long-range restraints, could be obtained from 3D ^{15}N-separated, ^{13}C-separated, and ^{13}C/^{13}C-separated NOESY spectra, and were used in the structure calculation. Thus, the results demonstrated that CYBAY Bayesian-assisted structure refinement along with QME data processing and FLYA chemical shift assignment is a powerful method for the in-cell structure determinations even without using hydrogen bond restraints.[10]

The Bayesian-assisted structure refinement was also employed for the structure determinations in sf9 cells. For the GB1 structure determination in sf9, approximately 98% of the backbone resonances were unambiguously assigned owing to high-quality 3D triple-resonance NMR spectra. Thus, Bayesian inference-assisted structure refinement based exclusively on in-cell data yielded a well-converged structure ensemble of GB1 with an average backbone RMSD of 0.51 Å to the mean coordinates (Figure 5.7).

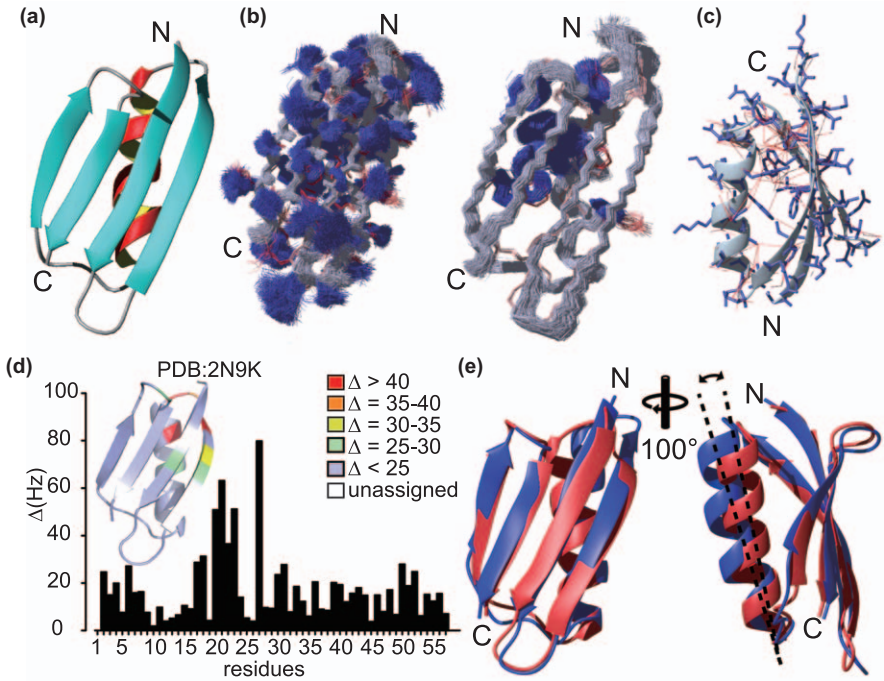


Figure 5.7 NMR structure of the protein GB1 in living sf9 cells. (a) The structure of GB1 in living sf9 cells with the highest posterior probability density in the Bayesian inference calculation. (b) Backbone heavy atoms of the structure ensemble of GB1 in sf9 cells (grey) superimposed onto the 20 structures in diluted solution (red). Side-chain (left) and aromatic residues (right) are highlighted with blue. (c) Distance restraints (red) shown with side-chains. (d) Chemical shift differences for GB1 in sf9 cells and diluted solution. (e) Superposition of the structures in diluted solution (lowest energy; red) and the sf9 (highest posterior; blue).
Adapted from ref. 11 with permission from John Wiley and Sons, Copyright © 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

Structures of Ub3A and TTHA1718 were also determined with distance restraints derived from 3D NOESY spectra in sf9 cells. The chemical shift assignments for these proteins were directly transferred from the data in diluted solution based on the observation that their chemical shift differences were small between sf9 cells and diluted solution. The structure ensemble of Ub3A was well-converged with an average backbone RMSD of 0.39 Å to the mean coordinates, and 1.31 Å to that in diluted solution (Figure 5.8a). The structure ensemble of TTHA1718 presented an average backbone RMSD of 0.88 Å to the mean coordinates and 2.60 Å to that in diluted solution, respectively (Figure 5.8b). As in the case of the structure determination in *E. coli*, the relatively large RMSDs of TTHA1718 were attributed to the putative metal-binding loop region of residues 9–18, for which only few NOE distance restraints were collected, presumably due to

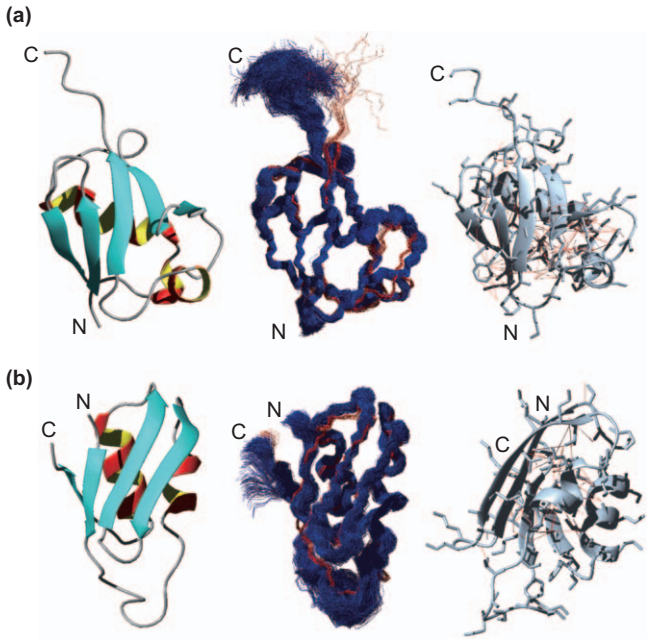


Figure 5.8 Protein structures of Ub3A and TTHA1718 in living sf9 cells. (a) Ribbon diagram of the Ub3A structure in sf9 cells with the highest posterior (left). Ub3A structures in sf9 cells (blue) and in diluted solution (red), showing the backbone (N, C^{α}, C′) atoms (centre). Distance restraints (red) for the structure calculation of Ub3A, shown in a ribbon model with side-chains (right). (b) Same as (a) for TTHA1718.
(a) and (b) Adapted from ref. 11 with permission from John Wiley and Sons, Copyright © 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

exchange processes related to the binding of various metal ions.[11] Excluding this region, the backbone RMSD to the structure in diluted solution was 1.27 Å. These results demonstrate again the high versatility of this approach with Bayesian-assisted structure refinement, allowing to obtain accurate 3D structures of proteins of less than 10 kDa size in sf9 cells.

5.4 Protein 3D Structures in an Intracellular Environment

For the first in-cell structure determination of TTHA1718, the structure of TTHA1718 in *E. coli* cells agreed well with the one determined *in vitro*, showing 1.16 Å of backbone RMSD. Slight structural differences were observed in some loop regions which may have high mobility and be susceptible to the effects of viscosity and molecular crowding in the cytosol. The most significant difference was found in the putative heavy metal-binding loop that corresponds to the region having relatively large chemical shift differences. Interactions with metal ions in the *E. coli* cytosol might affect

the conformation. Indeed, for the C11S/C14S and C11A/C14A mutants, which lack metal-binding activity, no significant chemical shift changes in *E. coli* cells were observed even when the cells were loaded with excess metal ions such as that for the wild-type protein.[7]

For the GB1 structure in *E. coli* cells, highly accurate ensemble structures were obtained by the three methodological advances described above with a backbone RMSD to the *in vitro* structure of 1.18 Å.[10] RMSDs of C^{α} atoms were below 1.0 Å for most residues, except for two loops of residues 22 and 50–51 that showed slightly higher values around 1.0 Å. A loop and the end of a β-strand (residues 11–14) showed low RMSDs to the *in vitro* structure for the C^{α} atoms but higher RMSDs of more than 2.0 Å for the side-chains. These residues coincided with a region of slightly higher chemical shift differences between the in-cell and *in vitro* samples (residues 10–13). As is the case with TTHA1718, molecular crowding effects or the intracellular environment might contribute to this slight structural change to the side-chains which is probably due to non-specific charge–charge interactions. It is known that most proteins in *E. coli* are polyanions at physiological conditions. Considering that these residues are on the molecular surface and include two lysines, interactions with other negatively charged molecules might result in structural changes to side-chains.

In living sf9 cells, the GB1 structure also agreed well with that in diluted solution with a backbone RMSD of 1.61 Å, except for a region composed of a loop and an α-helix (residues 22–26, 28) with higher RMSD values around 1.5 Å (Figure 5.7e). These residues coincided well with a region exhibiting chemical shift differences between the in-cell and diluted solution samples (residues 20–24, 27; Figure 5.7d). The relative position of the α-helix in the Sf9 structures was tilted away from the β-sheet (Figure 5.7e). This structural difference was very similar to the one observed in the GB1 structure in *E. coli* cells (residues 20–24), which was also supported by a molecular dynamics study in crowded environments[45] in that the intracellular environment perturbed the conformation of the region similarly in *E. coli* and Sf9 cells. The authors concluded that the changes in chemical shifts and 3D structure for this region can also be attributed to non-specific interactions through a hydrophobic patch on the protein surface with other molecules in the intracellular environment. Meanwhile, GB1 is currently the only protein whose structure has been determined very accurately in living cells. As future work, it would be necessary to validate the in-cell structures in various ways, such as by measuring paramagnetic effects and by observing structural dynamics in cells with NMR relaxation experiments. Verifications of this finding for other proteins in different intracellular environments or artificial crowders may also be needed.

5.5 Perspectives

In this chapter, we focused on *de novo* in-cell structure determination mostly based on distance restraints derived from 3D NOESY type experiments, along with advanced methodologies so as to overcome the short

lifetime and low signal sensitivity of in-cell spectra. Using the methods described here, one can currently achieve high-resolution 3D spectra and structures for small-size proteins with molecular weights of approximately 10 kDa, even in the limited measurement time and sparse experimental data in living systems. However, it is still not trivial to determine the structures of proteins beyond 10 kDa size, and hence further methodological improvements will be needed. Paramagnetic effects such as paramagnetic relaxation enhancement (PRE), pseudo-contact shift (PCS), and residual dipolar coupling (RDC)[46–48] may be promising candidates to elucidate 3D structures of proteins. As general issues for the observation of paramagnetic effects, the method requires to incorporate a paramagnetic probe which must be chemically and dynamically stable with a short covalent linker to the protein for collecting accurate structural information. In addition, in-cell NMR experiments cannot utilize probes with disulphide bonds, which are often used in chemical modifications to proteins, because of the intracellular reducing environment. Meanwhile, lanthanide-binding or nitroxide radical tags recently developed make it possible to measure PCS and PRE in the reducing conditions.[49–53] These new stable tags in the reducing environments inside cells would be promising candidates for obtaining accurate conformational data in future in-cell structure determination. New developments of the bioreactor system will also contribute to maintaining the health of cells and proteins, such as minimizing the loss of signal sensitivity due to the suppression of cell mobility and inhomogeneity in an NMR sample tube.[54] Moreover, stereospecific isotope labelling[55,56] and segmental labelling techniques[57–61] would be valuable for future applications of in-cell NMR. State-of-the-art signal reconstruction methods such as compressed sensing[28,29,62,63] may be able to extract further structural information from in-cell data. Over time these methods will jointly help to elucidate the fundamental mechanisms and unknown factors in living systems.

Acknowledgements

We gratefully acknowledge financial support from the Funding Program for Core Research for Evolutional Science and Technology (CREST; JPMJCR13M3) from the Japan Science and Technology Agency (JST), Next Generation World-Leading Researchers (NEXT Program), Grants-in-Aid for Scientific Research (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, JP17H05887 and JP19H05773 to Y.I.) from the Japan Society for the Promotion of Science (JSPS).

References

1. R. J. Ellis, *Trends Biochem. Sci.*, 2001, **26**, 597–604.
2. Z. Serber, A. T. Keatinge-Clay, R. Ledwidge, A. E. Kelly, S. M. Miller and V. Dötsch, *J. Am. Chem. Soc.*, 2001, **123**, 2446–2447.

3. E. Luchinat and L. Banci, *J. Biol. Chem.*, 2016, **291**, 3776–3784.
4. T. Ikeya, D. Ban, D. Lee, Y. Ito, K. Kato and C. Griesinger, *Biochim. Biophys. Acta, Gen. Subj.*, 2018, **1862**, 287–306.
5. S. Dzatko, M. Krafcikova, R. Hänsel-Hertsch, T. Fessl, R. Fiala, T. Loja, D. Krafcik, J. L. Mergny, S. Foldynova-Trantirkova and L. Trantirek, *Angew. Chem., Int. Ed.*, 2018, **57**, 2165–2169.
6. Z. Serber and V. Dötsch, *Biochemistry*, 2001, **40**, 14317–14323.
7. D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Wälchli, B. O. Smith, M. Shirakawa, P. Güntert and Y. Ito, *Nature*, 2009, **458**, 102–105.
8. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
9. Y. Ito and T. Ikeya, *Experimental Approaches of NMR Spectroscopy – Methodology and Application to Life Science and Materials Science*, ed. A. Naito, The Nuclear Magnetic Resonance Society of Japan, Springer, Singapore, 2017, ch. 3, pp. 63–90.
10. T. Ikeya, T. Hanashima, S. Hosoya, M. Shimazaki, S. Ikeda, M. Mishima, P. Güntert and Y. Ito, *Sci. Rep.*, 2016, **6**, 38312.
11. T. Tanaka, T. Ikeya, H. Kamoshida, Y. Suemoto, M. Mishima, M. Shirakawa, P. Güntert and Y. Ito, *Angew. Chem., Int. Ed.*, 2019, **58**, 7284–7288.
12. P. Selenko, Z. Serber, B. Gadea, J. Ruderman and G. Wagner, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11904–11909.
13. J. Hamatsu, D. O'Donovan, T. Tanaka, T. Shirai, Y. Hourai, T. Mikawa, T. Ikeya, M. Mishima, W. Boucher, B. O. Smith, E. D. Laue, M. Shirakawa and Y. Ito, *J. Am. Chem. Soc.*, 2013, **135**, 1688–1691.
14. N. G. Sharaf, C. O. Barnes, L. M. Charlton, G. B. Young and G. J. Pielak, *J. Magn. Reson.*, 2010, **202**, 140–146.
15. S. Kubo, N. Nishida, Y. Udagawa, O. Takarada, S. Ogino and I. Shimada, *Angew. Chem., Int. Ed. Engl.*, 2013, **52**, 1208–1211.
16. T. Ikeya, A. Sasaki, D. Sakakibara, Y. Shigemitsu, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, D. Nietlispach, M. Wälchli, B. O. Smith, M. Shirakawa, P. Güntert and Y. Ito, *Nat. Protoc.*, 2010, **5**, 1051–1060.
17. M. K. Rosen, K. H. Gardner, R. C. Willis, W. E. Parris, T. Pawson and L. E. Kay, *J. Mol. Biol.*, 1996, **263**, 627–636.
18. Ē. Kupče and R. Freeman, *J. Magn. Reson.*, 2003, **163**, 56–63.
19. S. G. Hyberts, H. Arthanari, S. A. Robson and G. Wagner, *J. Magn. Reson.*, 2014, **241**, 60–73.
20. Ē. Kupče and R. Freeman, *J. Am. Chem. Soc.*, 2004, **126**, 6429–6440.
21. P. N. Reardon and L. D. Spicer, *J. Am. Chem. Soc.*, 2005, **127**, 10848–10849.
22. J. Barna, E. Laue, M. Mayger, J. Skilling and S. Worrall, *J. Magn. Reson.*, 1969, **73**, 69–77.
23. P. Schmieder, A. S. Stern, G. Wagner and J. C. Hoch, *J. Biomol. NMR*, 1994, **4**, 483–490.

24. D. Rovnyak, D. P. Frueh, M. Sastry, Z.-Y. J. Sun, A. S. Stern, J. C. Hoch and G. Wagner, *J. Magn. Reson.*, 2004, **170**, 15–21.
25. D. Rovnyak, J. C. Hoch, A. S. Stern and G. Wagner, *J. Biomol. NMR*, 2004, **30**, 1–10.
26. W. Boucher, University of Cambridge, unpublished.
27. K. Kazimierczuk and V. Y. Orekhov, *Angew. Chem., Int. Ed. Engl.*, 2011, **50**, 5556–5559.
28. D. J. Holland, M. J. Bostock, L. F. Gladden and D. Nietlispach, *Angew. Chem., Int. Ed. Engl.*, 2011, **50**, 6548–6551.
29. Y. Shigemitsu, T. Ikeya, A. Yamamoto, Y. Tsuchie, M. Mishima, B. O. Smith, P. Güntert and Y. Ito, *Biochem. Biophys. Res. Commun.*, 2015, **457**, 200–205.
30. P. Güntert and L. Buchner, *J. Biomol. NMR*, 2015, **62**, 453–471.
31. E. Schmidt and P. Güntert, *J. Am. Chem. Soc.*, 2012, **134**, 12817–12829.
32. T. Ikeya, J. G. Jee, Y. Shigemitsu, J. Hamatsu, M. Mishima, Y. Ito, M. Kainosho and P. Güntert, *J. Biomol. NMR*, 2011, **50**, 137–146.
33. E. Schmidt and P. Güntert, *J. Biomol. NMR*, 2013, **57**, 193–204.
34. G. Cornilescu, F. Delaglio and A. Bax, *J. Biomol. NMR*, 1999, **13**, 289–302.
35. P. Güntert, C. Mumenthaler and K. Wüthrich, *J. Mol. Biol.*, 1997, **273**, 283–298.
36. A. Jain, N. Vaidehi and G. Rodriguez, *J. Comput. Phys.*, 1993, **106**, 258–268.
37. M. Habeck, M. Nilges and W. Rieping, *Phys. Rev. E: Stat. Nonlin. Soft Matter Phys.*, 2005, **72**, 031912.
38. W. Rieping, M. Habeck and M. Nilges, *Science*, 2005, **309**, 303–306.
39. R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.*, 1986, **57**, 2607–2609.
40. K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.*, 1996, **65**, 1604–1608.
41. S. A. Shahid, B. Bardiaux, W. T. Franks, L. Krabben, M. Habeck, B. J. van Rossum and D. Linke, *Nat. Methods*, 2012, **9**, 1212–1217.
42. T. Ikeya, S. Ikeda, T. Kigawa, Y. Ito and P. Güntert, *J. Phys.: Conf. Ser.*, 2016, **699**, 012005.
43. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.
44. N. A. Baker, *Curr. Opin. Struct. Biol.*, 2005, **15**, 137–143.
45. R. Harada, N. Tochio, T. Kigawa, Y. Sugita and M. Feig, *J. Am. Chem. Soc.*, 2013, **135**, 3696–3701.
46. G. M. Clore and J. Iwahara, *Chem. Rev.*, 2009, **109**, 4108–4139.
47. G. Otting, *Annu. Rev. Biophys.*, 2010, **39**, 387–405.
48. T. Saio and F. Inagaki, *Experimental Approaches of NMR Spectroscopy – Methodology and Application to Life Science and Materials Science*, ed. A. Naito, The Nuclear Magnetic Resonance Society of Japan, Springer, Singapore, 2017, ch. 8, pp. 227–252.
49. T. Müntener, D. Häussinger, P. Selenko and F. X. Theillet, *J. Phys. Chem. Lett.*, 2016, **7**, 2821–2825.

50. B. B. Pan, F. Yang, Y. Ye, Q. Wu, C. Li, T. Huber and X. C. Su, *Chem. Commun.*, 2016, **52**, 10237–10240.
51. Y. Hikone, G. Hirai, M. Mishima, K. Inomata, T. Ikeya, S. Arai, M. Shirakawa, M. Sodeoka and Y. Ito, *J. Biomol. NMR*, 2016, **66**, 99–110.
52. T. Müntener, J. Kottelat, A. Huber and D. Häussinger, *Bioconjug Chem*, 2018, **29**, 3344–3351.
53. Y. Yang, J. T. Wang, Y. Y. Pei and X. C. Su, *Chem. Commun.*, 2015, **51**, 2824–2827.
54. K. Inomata, H. Kamoshida, M. Ikari, Y. Ito and T. Kigawa, *Chem. Commun.*, 2017, **53**, 11245–11248.
55. M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. Mei Ono and P. Güntert, *Nature*, 2006, **440**, 52–57.
56. M. Kainosho, Y. Miyanoiri, T. Terauchi and M. Takeda, *J. Biomol. NMR*, 2018, **71**, 119–127.
57. D. Liu, R. Xu and D. Cowburn, *Methods Enzymol.*, 2009, **462**, 151–175.
58. J. Xue, D. S. Burz and A. Shekhtman, *Adv. Exp. Med. Biol.*, 2012, **992**, 17–33.
59. Y. Minato, T. Ueda, A. Machiyama, I. Shimada and H. Iwai, *J. Biomol. NMR*, 2012, **53**, 191–207.
60. L. Freiburger, M. Sonntag, J. Hennig, J. Li, P. Zou and M. Sattler, *J. Biomol. NMR*, 2015, **63**, 1–8.
61. K. M. Mikula, I. Tascon, J. J. Tommila and H. Iwai, *FEBS Lett.*, 2017, **591**, 1285–1294.
62. S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari and G. Wagner, *J. Biomol. NMR*, 2012, **52**, 315–327.
63. J. Ying, F. Delaglio, D. A. Torchia and A. Bax, *J. Biomol. NMR*, 2017, **68**, 101–118.