Mini Review

Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins
Open Avenues for Life Scientists

Toshihiko Sugiki *, Naohiro Kobayashi, Toshimichi Fujiwara

Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for structural studies of chemical compounds and biomolecules such as DNA and proteins. Since the NMR signal sensitively reflects the chemical environment and the dynamics of a nuclear spin, NMR experiments provide a wealth of structural and dynamic information about the molecule of interest at atomic resolution. In general, structural biology studies using NMR spectroscopy still require a reasonable understanding of the theory behind the technique and experience on how to recorded NMR data. Owing to the remarkable progress in the past decade, we can easily access suitable and popular analytical resources for NMR structure determination of proteins with high accuracy. Here, we describe the practical aspects, workflow and key points of modern NMR techniques used for solution structure determination of proteins. This review should aid NMR specialists aiming to develop new methods that accelerate the structure determination process, and open avenues for non-specialist and life scientists interested in using NMR spectroscopy to solve protein structures.

© 2017 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords:
Nuclear magnetic resonance (NMR) spectroscopy
Solution NMR
Automation
Protein Structure determination Validation

Contents

1. Introduction .............................................................. 328
2. Workflow of Protein Structure Determination by Solution NMR ................................................................. 329
3. Protein Sample Preparation for NMR Measurements .......................................................................................... 330
4. Acquisition of NMR Spectra and Their Assignment ............................................................................................ 331
5. NMR Structure Calculation ............................................. 332
   5.1. Distance Restraints ................................................... 332
   5.2. Dihedral Angle Restraints ........................................... 333
6. Automated NOE Assignment and Structure Modeling of Proteins ................................................................. 333
7. Validation of Precision and Accuracy of the Determined Protein Structure. ......................................................... 334
8. Concluding Remarks........................................................ 335

Acknowledgements ........................................................................... 335
Appendix A. Supplementary Data .................................................. 335
References ................................................................................... 335

1. Introduction

Tertiary structure determination of biomolecules such as nucleic acids and proteins at atomic resolution provides essential insight into the function of bioactive molecules. X-ray crystallography [1] and nuclear magnetic resonance (NMR) spectroscopy [2] have been the primary methods over the past few decades to obtain high-resolution structures. More recently, the rapid technological growth of cryo-electron microscopy [3] has seen this technique emerge as a third major approach to solve biomacromolecular structures at atomic resolution. Solution NMR offers a number of distinct features for structural biology studies: 1) Dynamics of protein folding, structural fluctuations, internal mobility...
and chemical exchange of target molecules can be investigated over a wide range of timescales (i.e., between picoseconds and sub-millisecond timescales) [4,5]. The most successful application of NMR to study biomolecular dynamics has been carried out on intrinsic disordered proteins (IDPs) such as zinc finger proteins that usually do not yield crystals, which provide sufficient quality X-ray diffraction data for high-resolution structure determination [6–11]. Moreover, these proteins are too small or too flexible to obtain strong contrast images by modern cryo-EM analysis. 2) Studies of protein-protein or protein–ligand interactions can be performed under physiological conditions. The affinity and the location of the interaction sites between the target protein and its binding partner molecules can be determined accurately and sensitively even if the interaction is very weak (K_a = -nM) by performing simple NMR experiments, e.g., chemical shift perturbation analysis by measuring two-dimensional (2D) heteronuclear single-quantum coherence (HSQC) spectra of target proteins with and without the binding partner, while such an interaction study by co-crystallization or co-immunoprecipitation can be challenging [12–14]. Therefore, techniques based on using solution NMR can be readily used to screen chemical fragment libraries for preliminary hits, which have the potential to act as seed compounds in drug development. Such approaches usually identify weak binders for the target protein, but represent a powerful starting point for structure-based drug discovery and development studies.

Among the various applications of solution NMR to study protein function, as aforementioned, tertiary structure determination of proteins by NMR remains a steadfast and powerful application of this spectroscopy. Recent development of software and a number of web-based resources, which reduce the burden of complex NMR data analysis, have contributed to the systematic integration of sophisticated semi-automated NMR platforms for structure determination of biomolecules [15–19]. Such advances, along with improvements in NMR hardware, have lowered the knowledge barrier to facilitate entry into the field of structural biology by NMR for non-specialists. However, only a handful of articles that describe the workflow and practical aspects of protein structure determination by solution NMR spectroscopy, which covers the latest method developments, have been published despite the practicality of this technique.

2. Workflow of Protein Structure Determination by Solution NMR

The standard approach to protein structure determination by solution NMR comprises several steps: (i) preparation of a protein sample labeled with stable NMR spin-1/2 isotopes, ^13C and ^15N; (ii) acquisition of NMR spectra; (iii) data processing and assignments of signals; (iv) generation of a chemical shift table that gives the signal assignments derived from the analysis of NMR spectra such that the analyst or an automated program can subsequently assign NOE signals to generate

![Diagram](http://bmrbdrep.pdbj.org/en/nmr_tool_box/magro_nmrview.html)

Fig. 1. Workflow of protein structure determination by solution NMR spectroscopy. (A) First, the target protein is uniformly labeled with NMR active isotopes (^13C and ^15N) and purified by chromatography processes. In many cases, it is essential to optimize the solution composition of the NMR sample (e.g., buffer, pH, type and concentration of salt, and other additives to prevent aggregation and/or to improve thermal stability of the target protein) and NMR parameters (e.g., sample temperature during NMR data collection, strength of static magnetic field, pulse sequences, inversion recovery delay) prior to collecting all the multidimensional heteronuclear NMR experiments. (B) Generally, various sample conditions and NMR parameters are verified by measuring 2D ^1H–^15N HSQC NMR spectra and/or 2D projection spectra of 3D HNCACB (e.g., 2D HN(CA)CB) until optimal conditions are found. If sample/NMR data collection optimization is performed by only ^15N-based NMR experiments, only a ^15N-labeled target protein (without ^13C-labeling) is required. (C) The NMR signals of all ^1H, ^13C and ^15N nuclei are assigned by analyzing multidimensional heteronuclear NMR spectra using NMR software that facilitates assignment of the signals. The right lower panel of this figure is a screenshot of a NMR signal assignment using the Kujira and MagRO software, developed by Prof. Naohiro Kobayashi (Osaka University, Japan), which is available at PDBj-BMRB http://bmrbdrep.pdbj.org/en/nmr_tool_box/magro_nmrview.html. (D) After completing the assignment process, the solution structures of the target protein are determined by performing simulated annealing using NMR-based restraints such as ^1H–^1H distances and dihedral angles derived from NOEs and chemical shifts, respectively. (E) Finally, the precision and the accuracy of the determined structure are verified statistically and experimentally (e.g., by analyzing the Ramachandran plot and measuring RDCs of the target protein, respectively).
distance or dihedral angle restraints; and (v) simulated annealing (SA) by simplified molecular dynamics calculations with the NMR-based restraints are used for initial structure modeling to obtain an ensemble of structures that satisfy the experimentally determined constraints. These SA calculations are followed by molecular dynamics (MD) simulation with explicit or implicit water using a more sophisticated force field. (vi) The final stage of the workflow is validation of the precision and accuracy of the determined structure (Fig. 1). Details of these steps are explained in the subsections below. In this mini-review, basic and general information for researchers interested in NMR structure determination of proteins will be provided. We will mainly describe general procedures that are routinely used for NMR structure determination of soluble and small-to-medium size (c.a. <25 kDa) proteins. More advanced techniques required for NMR analysis of more challenging proteins, e.g., membrane proteins or larger proteins (c.a. >25 kDa), will also be described briefly.

3. Protein Sample Preparation for NMR Measurements

In the initial NMR studies aimed at solving protein structures, complete assignment of the proton (1H) NMR signals was achieved by acquiring 1H NMR spectra, e.g., 2D 1H-1H double quantum filtered-correlation spectroscopy (DQF-COSY), 2D 1H-1H totally correlation spectroscopy (TOCSY), and 2D 1H-1H NOE correlated spectroscopy (NOESY) spectra. These spectra provided sufficient information to obtain 1H-1H distance restraints to calculate the tertiary structure for non-isotope labeled peptides and small proteins (c.a. <10 kDa) [20, 21]. Extension of the 1H only approach to larger proteins was not possible because of severe signal overlap precluding unambiguous assignment of signals [21]. In the early 1990s, owing to the development of technologies for expressing recombinant proteins using bacteria cultures with medium containing stable isotopes, a large number of target proteins could be uniformly labeled with the stable "NMR active" isotopes, 13C and 15N, which can be detected by NMR measurements. Two- (2D), three- (3D) and four- (4D) dimensional NMR spectra observing signals that correlated 1H, 13C and 15N nuclei made for dramatic improvements in signal separation and removed signal overlap present in 1H NMR spectra [21].

The Escherichia coli (E. coli) expression system is most commonly used for stable isotope labeling of proteins, as it has several advantages: easy handling, bacteria cells grow quickly, cost efficiency and the availability of many established methods for protein isotope labeling [22,23]. For example, the most powerful application of the methods would be site-specific/desired amino acid selective isotope labeling of proteins, which has been firmly established using the E. coli expression system [22–25]. Alternative 15C-labeling of protein is also possible by using particular 15C-enriched carbon sources, such as [1-13C]-glucose [26], [1,3-13C2]- or [2-13C]-glycerol [27], 13C-acetate [28], and [1,2-13C2]- or [3-13C]-pyruvate [29–31]. As this technique enables NMR observation of only the desired moiety of the target protein, it is especially useful for large molecular-weight (c.a. >25 kDa) proteins or IDPs that show severe signal degeneracy.

In situations where an insufficient amount of the recombinant protein is produced by the E. coli expression system, the yeast expression systems, Pichia pastoris [32] or Klyuveromyces lactis [33] as host cells, can be used as an alternative. These systems not only have the advantage of being similar to the E. coli expression system (easy handling, rapid cell growth, and many available and firmly established isotope labeling techniques with relatively low cost), but also offer eukaryotic cell-specific features that facilitate the over-expression and correct folding of particular proteins, e.g., protein requires a complex array of disulfide bonds to form for native folding [22,23,34,35].

Cell-free protein expression systems are another option to overexpress proteins. This approach synthesizes recombinant proteins in vitro in a reaction mixture containing the DNA coding the target protein and a cell lysate containing gene transcription/translation machinery, and, if required, isoence labeled amino acids for protein labeling [36–40]. Preparation of a fresh cell lysate, which is necessary for synthesizing a sufficient amount of the recombinant protein stably and reproducibly, is challenging. Furthermore, the running cost of such a cell-free system is generally high [23], indicating that it is not suitable for large-scale protein production with stable isotopes. However, this approach is still adequate for amino acid selective isotope labeling or selective introduction of unnatural amino acid residues into the target protein, because amino acid scrambling is relatively mild when compared with other expression systems that use host cells [41–46]. As an additional remarkable point, reconstructing an isotope-labeled membrane protein into artificial lipid bilayer discs, e.g., nanodiscs, is relatively easy to achieve by cell-free expression [47–50].

Isotope labeling of recombinant protein using insect or mammalian cells is also possible, although this is generally an expensive method with low yields. Moreover, the available isotope labeling variation is limited when compared with that of other expression systems [51,52].

In NMR protein structure determination, sample purity should be as high as possible because signal linewidths and protein stability are affected by non-specific contact with impurities. However, NMR spectra can be obtained even if the sample is crude, as demonstrated by studies examining cell lysates or metabolites [53–58].

The conditions of the protein solution and the NMR experimental parameters should be sufficiently optimized prior to starting multidimensional heteronuclear NMR measurements [59–61]. In many cases, the protein concentration, hydrogen ion (pH) value of the sample solution, the type of salt/buffer and its concentration and the NMR measurement temperature should be carefully examined by several pilot NMR experiments (typically 1D 1H or 2D 1H-15N HSQC spectra are recorded) with different sample solution compositions and NMR parameters. If an optimal condition is found, the NMR spectra will show sharp (narrow linewidths) and sufficiently well-dispersed signals [59,61,62]. To assign backbone 1H, 13C and 15N signals successfully by the inter-residue chemical shift linking approach, the observation of 13C nuclei with reasonable chemical shift dispersion and sensitivity is a key point to consider. Pilot measurements of 3D HNCA CB and/or 2D HN(CA)CB spectra can be good indicators for judging the adequacy of the solution conditions and potential difficulties of obtaining signal assignments. Perdeuteration of target proteins combined with transverse relaxation optimized spectroscopy (TROSY)-type NMR pulse schemes may be beneficial when the molecular weight of the target protein exceeds 25 kDa (see Section 4).

Although the sensitivity of solution NMR is continuously improving thanks to the development of new NMR pulse schemes and hardware, e.g., an increase in the static magnetic field strength and the innovation of cryogenic probe [63], a higher protein concentration is preferable to obtain NMR signals with higher signal-to-noise ratio and resolution within a reasonable period of NMR instrument use. In particular, for protein structure determination, sufficiently high solubility and stability are typically required. Ideally, the target protein should be a monodispersed species at a protein concentration ~0.5 mM and be stable in the NMR instrument for a minimum of 1–5 days.

The aggregation character of the target protein can be monitored by observing broadening of the NMR signals as a function of protein concentration or by investigating the rotational correlation time of the target protein via measurement of the 1H-1H-1H-1H-TRACT [64] or the 1H2N T1/T2 value [65]. Biochemical wet experiments (e.g., size-exclusion column chromatography) are also useful for investigating the character of the target protein. However, this method cannot address protein aggregation at the protein concentration used for NMR measurements, because the protein is significantly diluted in chromatography [59].

In general, the pH value of the protein solution is selected by considering the theoretically/experimentally determined isoelectric point (pI) of the protein. A lower pH is preferable to reduce the chemical exchange between water 1H and protein 1H5 nuclei, because 1H fast exchange
reduces the sensitivity of $^1$H$^N$-$^1$N correlation signals [66,67]. In some cases, however, the protein cannot be placed at a low pH because of the pl, solubility, or stability. As buffer components, phosphate buffer (e.g., sodium or potassium phosphate) is an ideal choice because it does not have $^1$H nuclei [61]. As a character of buffer conductivity and its effect on the sensitivity of NMR signal detection using cryogenic probes, however, other buffers such as HEPES- or MES-NaOH would be more preferable [68].

The sample temperature during NMR data collection is often room temperature (typically 25 °C). However, a higher sample temperature should give better spectra unless such temperatures affect the structure and stability of the target protein (up to 40 °C when a cryogenic probe is used). Higher temperatures offer sharper signals by decreasing solvent viscosity and thus the rotational correlation time of the protein [59,61].

Some types of chemical additives can increase the solubility and/or stability of proteins [59,61]. For example, a reducing agent (e.g., dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP)-HCl) may be essential for preventing protein oligomerization caused by inter-molecular nonspecific disulfide bond formation, when free cysteines exist on the surface of the protein. Multiple well plate-based high-throughput screening methods using fluorescence correlation spectroscopy or thermal shift assays can explore chemical additives that improve protein solubility and stability, which may help to find the best condition without exhaustive sample condition optimization work that consumes limited NMR instrument availability [59,61].

Note that 5%-10% (v/v) $^2$H$_2$O must be added to the NMR sample to provide a lock signal for the static magnetic field passing through the sample solution. In addition, in many cases, a standard chemical shift reference compound is dissolved in the sample solution, as described in the next section.

4. Acquisition of NMR Spectra and Their Assignment

Initially, $^1$H, $^{13}$C, and $^{15}$N NMR signals arising from the nuclei of the target protein have to be assigned before we can obtain angle and distance restraints for protein solution structure determination. In general, a number of NMR spectra are recorded to obtain a complete set of NMR assignments [69–71].

For backbone ($^1$H, $^{13}$C and $^{15}$N) signal assignments: 2D $^1$H-$^{15}$N HSQC, 3D HNCA, $^3$D HN(COA)’, 3D HNCA/CB, 3D CBCA(CO)/NH, 3D HNCO, 3D HN(CA)CO’ and 3D HBHA(CO)/NH are recorded. The 3D HNCA/CB and 3D CBCA(CO)/NH are the most basic spectra for backbone signal assignments [71]. The asterisk next to the experiment acronym indicates optional spectra required when 3D HNCA/CB and 3D CBCA(CO)/NH spectra are sufficient for completing backbone assignment. The 3D HNCO and 3D HN(CA)CO are useful complements to inter-residue chemical shift linking using 3D HNCA/CB and 3D CBCA(CO)/NH experiments, and also help to eliminate any potential assignment ambiguity that exists. Furthermore, chemical shift assignments of $^{13}$Cα, $^{13}$Cβ, $^{13}$Cγ and $^1$Hα nuclei are important for secondary structure prediction by TALOS or its related programs (e.g., TALOS+, TALOS-N) as described in Section 5.2. In those pulse schemes, except CBCA(CO)/NH, $^1$H–$^{13}$C–$^{15}$N correlations are measured using the out-and-back coherence transfer approach between $^1$H and $^{13}$C/$^{15}$N to improve the sensitivity of those low gyromagnetic ratio ($\gamma$) nuclei (typically 5–10 fold). First, polarization of proton magnetization is transferred to low-sensitive nuclei $^{13}$C or $^{15}$N via f-couplings, and then the chemical shifts of those nuclei are recorded. Finally, the magnetization of $^{13}$C or $^{15}$N is returned to the starting proton, and the chemical shifts of the $^1$H nuclei are recorded as the free induction decay (FID). However, nuclei denoted within brackets in the acronym of the pulse sequence participate in the coherence transfer pathway but their chemical shifts are not encoded in those multi-dimensional NMR spectra. Incidentally, those out-and-back coherence transfer schemes starting from the amide proton are also applicable to perdeuterated proteins.

For side chain assignments: 2D $^1$H–$^{13}$C constant-time (CT) HSQC ($^{13}$C offset on aliphatic and aromatic regions), 3D $^{15}$N-edited TOCSY-HSQC, 3D CC(CO)/NH, 3D H(C=CH)-TOSCY ($^{13}$C offset on aliphatic and aromatic regions), 3D H(C=CH)-TOCSY ($^{13}$C offset on aliphatic and aromatic regions), 2D (HB)CB(CGCD)HD, 2D (HB)CB(CGCDCE)HE, 13C-edited NOESY-HSQC and 15N-edited NOESY-HSQC spectra are required. There are many variations of those pulse schemes for time evolution such as CT or semi-CT methods, echo and anti-echo, single or multiple quantum coherence, or TROSY type acquisition. It is preferable that each user selects proper pulse schemes for individual protein samples by performing pilot NMR experiments to collect better quality NMR spectra that have the highest signal-to-noise ratio and resolution. Catalogues of standard pulse programs disclosed by the manufacturer of NMR instruments, such as Bruker (available from their websites), may help determine the best NMR measurements to use.

The minimum number of multi-dimensional NMR spectra required to obtain complete assignment of $^1$H, $^{13}$C, and $^{15}$N signals of proteins using magnetization transfer via inter-nuclear j-couplings is six: 3D HNCA/CB, 3D CBCA(CO)/NH, 3D H(C=CH)-TOSCY (aliphatic and aromatic regions) and 3D H(C=CH)-TOCSY (aliphatic and aromatic regions) [71]. In practice, however, using NOESY spectra (e.g., $^{13}$C-edited NOESY-HSQC), which provide a wide range of inter-nuclear distance information between the backbone and side chains, may be helpful, in particular, for assignment of aromatic side chain signals.

The $^{13}$C- and $^{15}$N-edited NOESY-HSQC spectra are essential for both signal assignments and for generation of $^1$H–$^1$H distance restraints [21,72]. The mixing time used in NOESY spectra of proteins is generally 80–150 ms, as described in Section 5.1.

In general, when the molecular weight of the target protein exceeds 30 kDa, measurement and assignment of the protein NMR signals become difficult owing to the increasing degeneration and line-broadening of the signals, of which effect is associated with fast transverse magnetization decay [21]. The latter issue is the biggest burden to solution NMR spectroscopy. For example, the transverse relaxation rate of $^{13}$Cα-magnetization is dominated by the dipole–dipole (DD) interaction with the proximate $^1$Hα, leading to dramatic sensitivity losses to $^{13}$Cα signals or signals in multidimensional NMR spectra that transfer magnetization via $^{13}$Cα. This is particularly problematic when the molecular weight of the target protein exceeds 20 kDa [73,74]. In this case, the DD relaxation of $^{13}$Cα magnetization can be suppressed remarkably by substitution of $^1$Hα to $^1$Hβ (perdeuteration) by overexpressing the target protein in $^2$H$_2$O media [73,74]. In addition, $^1$H–$^{13}$C correlation spectra with narrower signals can be obtained by TROSY-type NMR pulse schemes, which selectively observes the coherence where DD relaxation has been attenuated by the chemical shift anisotropy (CSA) effect [75]. The TROSY effect will be more prominent in combination with a perdeuterated protein and higher static magnetic fields (i.e., ≥ 800 MHz) because cancellation of DD relaxation by the CSA effect is field dependent. However, the benefits from perdeuterated proteins are limited to improving backbone NMR signal assignments and for $^{13}$C direct detection NMR experiments [76–79]. $^{13}$C direct detection NMR methods are also applicable to protein structure determination [80]. In many cases, moreover, mild denaturation and regeneration treatments of the deuterated protein sample in a $^1$H$_2$O buffer may be required to back-exchange deuterium to proton at amide positions to regenerate $^1$H–$^{13}$C–$^{15}$N correlations. However, multi-dimensional NMR experiments that use polarization transfer starting from non-labile protons, e.g. 3D CBCA(CO)/NH, cannot be used with deuterated proteins [71].

Prior to starting the assignment of $^1$H, $^{13}$C and $^{15}$N chemical shifts of the target protein, the NMR spectra should be calibrated [81,82]. $^1$H chemical shift calibration (or denoted as “referencing”) uses the temperature-dependency of the $^1$H chemical shift of $^2$H$_2$O automatically when processing NMR spectra using the program NMRPipe [83]. In addition, chemical shift calibration using standard compounds that yield $^1$H chemical shifts, e.g., 4,4-dimethyl-4-silapentane-1-sulfonic acid-d$_6$, T. Sugiki et al. / Computational and Structural Biotechnology Journal 15 (2017) 328–339
The basic concept of the automated signal assignment process is FLYA [98], which is partially powered by the GARANT algorithm [99] of MATCH/ASCAN/CANDID/ATNOS algorithms) [87,88,95].

Successful performance of automated assignment software requires the collection of high-quality NMR spectra that show sufficiently well-resolved signals and the expected number of signals. They typically provide reliable assignment of well-separated signals by linking chemical shifts, which are expected from coherence transfer pathways based on the pulse schemes used for the NMR experiments and topology of chemical structures of amino acid residues. Moreover, auto-assignment programs can offer possible assignment of unassigned signals, which are difficult to assign manually owing to ambiguity of inter-residue chemical shift linking. In many cases, therefore, the user has to manually confirm and carefully correct the assignment made by automated programs using NMR spectrum visualization software, as described above.

5. NMR Structure Calculation

5.1. Distance Restraints

Protein structure determination by simulated annealing uses the 1H–1H distance and dihedral angle restraints derived from NMR data [86,90]. The 1H–1H distance information is obtained by estimating the intensity of the 1H–1H NOE signals [21]: the intensity of the 1H–1H NOE signals is dependent on the mixing time (τm), during which 1H–1H cross-peaks are generated [21]. However, although build-up rates obtained from linear dependency of the NOE cross-peak intensity on the τm correlate with the 1H–1H distance when a limited (short) τm NOE build-up regime is used, quantitative estimation of 1H–1H distances from the intensities of the 1H–1H NOE signals becomes impossible over the limited τm range because relaxation and spin diffusion also proceed during long mixing times [21,100]. In general, the τm value should be within the initial NOE build-up range (typically 80–150 ms for proteins) [21]. Nonetheless, recently developed innovative methods offering more exact conversion of NOE data to intermolecular distance information, Exact NOE (eNOEs), are powerful protocols for accurate and precise protein solution structure determination [101–103]. The 13C- or 15N-edited NOESY-HSQC spectra are widely used. Following peak picking and integration of the 1H–1H NOE signals, the table containing the residue number/atom name/chemical shift/1H–1H NOE signal intensity is used as primary input data for simulated annealing [90]. In the classical way to prepare NOE based distance restraints, the NOE peaks are roughly classified into three (or four) groups depending on their signal intensity: strong, medium and weak (and very weak) [86,90]. In the structure calculation, this information is translated into 1H–1H distance ranges of 1.8–2.7, 1.8–3.3, and 1.8–5.0 (and 1.8–6.0) Å [86,90]. The lower limit of the 1H–1H distance, 1.8 Å, corresponds to twice the van der Waals radius of 1H atoms [86,90].

Rough classification and wide range (not rigorous) distance restrictions based on NOE data are based on the difficulty of converting the NOE signal intensity to an inter-nuclei distance accurately and directly because the NOE signal intensity is affected by not only the interatomic distance, but also various factors such as spin-diffusion and conformational averaging as a result of local structural fluctuations. Because of the mildness of distance restraints based on NOE experiments, a sufficient number of distance restraints are required for accurate and precise protein structure determination (generally speaking more than 10–20 distance restraints per residue, not including intra-residue NOEs) in order to build a sufficient density of NOE-network anchoring [90]. The concept of NOE-network anchoring is based on the consideration that the correctly assigned NOE distance restraints will form a self-consistent NOE set, which is compatible with tertiary structure models of target proteins [87,88]. It is also important to empower automated NOE assignment by the program CYANA [90] (see Section 6).

In particular, a sufficient number of long-range distance information (5.0–6.0 Å, which are upper limit distance restrictions) are critical for precise and accurate structure determination [104]. As the intensity of NOE
peaks with important distance information, i.e., long-range, is often weak, optimization of experimental conditions of the NOESY spectra is important to gather a sufficient number of weak NOE signals. Higher sample concentration (>0.5 mM), lower temperature, lower pH (in the case of 15N-edited NOESY-HSQC) and stronger static magnetic fields are preferable parameters to ensure observation of weak NOEs. When only a minimal number of NOE signals are observed owing to the poor character of the target protein, e.g., aggregation or signal line broadening owing to averaging of multiple conformation caused by fluctuation in an intermediate-slow NMR time scale regime, long-range distance information can be collected, which may complement the shortage of NOE-based distance restraints, by protein labeling with paramagnetic metal ions or radicals to measure the paramagnetic relaxation enhancement (PRE) effect [105] or the pseudo-contact shift (PCS) [106,107]. The PCS data provide important NMR-based restraints that describe the global fold of the protein and/or tertiary structure of protein–protein complexes with high accuracy owing to its unique features: PCS provides both long-range distance restraints and angle information of bond vectors in the protein against the tensor frame of the magnetically aligned paramagnetic center molecule.

In addition, the resolution of NOESY spectra significantly affects the quality of the structure determined by NMR [108]. Using the highest possible resolution is appropriate for precise and accurate structure determination; this is especially important when performing automated NOE peak picking and assignments. Algorithms for automated NOE peak picking, e.g., NMRView [91], AUTOPSY [109], ATNOS [88], AUDANA [110], or CYPICK [111], have been used for preparing peak lists to perform automated NMR analysis.

Information about the location of hydrogen bonds can be used as distance restraints [90]. The rate of chemical exchange between the amide proton in the protein and protons in the solvent water can be estimated by performing H/D exchange [112] or CLEANEX-PM [113] experiments. It is possible that amides protons showing a significantly slow exchange rate may be involved in the formation of hydrogen bonds. At a stage of the structure calculation process that the global fold of the protein has been trustfully determined, pair residues forming hydrogen bonds may be deduced from the modeled structure. However, this step must be carefully executed to prevent incorrect assignment of hydrogen bonded residues. More ideally, therefore, pair residues forming hydrogen bonds should be experimentally determined by long-range HNCO measurements [114,115] or other analogous NMR techniques to avoid incorrect interpretation of the modeled structure. The distances between backbone amide protons and carbonyl oxygen atoms, or backbone nitrogen and carbonyl oxygen, are normally set as 1.8–2.0 or 2.7–3.0 Å, respectively. Since the distance restraints of hydrogen bonds are strong, giving large entropic information that contrasts those of dihedral angles and NOE-based distance restraints, introduction of hydrogen bonds should be applied at a later stage in the structure calculation. In a similar manner, information of the location of intra-molecular disulfide bonds also available as restraints should be entered carefully into a structure calculation, perhaps after an initial model has been calculated using purely distance and dihedral restraints.

5.2. Dihedral Angle Restraints

Dihedral angles of the target protein, such as backbone ψ(N-Cα), ψ(Cα-C), ω(Nα-Cα-C) and side-chain χ(n), are determined by a wide variety of NMR experiments [86,90]. These angles are important for defining the secondary structure and the side-chain conformation of the target protein.

According to the Karplus equation, a dihedral angle can be estimated by measuring the 3J-coupling constant [116]. For example, the 3J-coupling constants of each residue can be directly measured by NMR experiments such as 3D HNHA [117], ARTSY-J [118] and 3D HNHB [119]. However, acquiring these experiments with high precision and accuracy cannot be achieved easily because the sensitivity of these NMR experiments that measure 3J-couplings is relatively low. In those pulse schemes, long evolution and refocusing periods (total ~50 ms in the case of 3D HNHA) are required to measure the rather small 3J-coupling constants (~< 10 Hz), which would be especially difficult to measure accurately with a protein that gives rise to broad signals owing to, for example, molecular size or chemical exchange processes [118]. As an alternative approach, dihedral angles and the secondary structure of proteins are predicted by the programs TALOS (and related versions) [84,85], CamShift [120] and SHIFTX2 [121]. These programs use prior chemical shifts statistics, including 1H, 15N, 13C, 13C, 1Hα and carbonyl 13C nuclei, with corresponding structure coordinates.

Dihedral angle information is useful to refine secondary structures in spite of their application in simulated annealing as relatively mild restraints (lower and upper limits are normally around +/−20–40°). For application of the predicted dihedral angles in structure modeling, it is important to use trusted chemical shifts derived from correctly calibrated NMR spectra, as described above.

6. Automated NOE Assignment and Structure Modeling of Proteins

Programs for automated NOE assignments and molecular modeling, e.g., CYANA [86,90,122], Xplor-NIH [123], AUREMOL (http://www.auremol.de/), critical assessment of automated structure determination of proteins by NMR (CASD-NMR) developed by the e-NMR project (http://haddock.chem.uu.nl/enmr/eNMR-portal.html) [19,124,125], PONDEROSA-C/S (http://p ponderosa.nmrfam.wisc.edu/) developed by the National Magnetic Resonance Facility at Madison (NMRFAM) (http://www.nmrfam.wisc.edu/software.htm) [16,18,126] and UNIO (http://perso.ens-lyon.fr/torsten.herrmann/Herrmann/Software.html) [19,87,88,95–97], are used for solution structure determination of biomolecules such as proteins and nucleic acids. In this mini-review, we will show an example of the program CYANA, which is used widely for NMR structure determination, to explain the general workflow of protein structure calculation. The amino acid sequence data of the target protein, the chemical shift table, and the NOE peak list are required as primary input files to perform automated NOE assignments and molecular modeling using CYANA [86,90,122]. In NOE based structure modeling by CYANA, seven steps of iterative calculations including the structure-modeling step for obtaining an initial global fold of the protein are performed. This is followed by refinements of the automated NOE assignments by referencing the calculated intermediate structure model of the protein using the algorithm extended from the program CANDID [90,122]. The automated NOE assignment process has the advantages of being fast, unambiguous and unbiased [88,127]. However, in many cases, manual intervention and/or re-assignment of the results of automated NOE assignment are necessary because completeness of the automated process is not perfect, especially when the peak maxima of the NOE signals do not match 1H chemical shift assignment table [88,127]. Several other algorithms, e.g., NOAH [128,129], ASDP [130], ARIA and its related programs [131,132], KNOWNOE [133], AutoNOE ROSETTA [134] and PASD in Xplor-NIH [135], have also been developed for automated NOE assignments and structure calculation. Even if the stereo-specific assignments of methylene protons or methyl groups of leucine or valine residues are ambiguous, in the process of the structure-aided NOE auto-assignment, those signals can be automatically swapped with fair correctness by accounting for the modeled structure just before the final stage in the calculation. A high precision of modeled structures can be strongly supported by random combination of long-range distance restraints and sufficiently condensed NOE network anchoring generated in the early stage of the CYANA calculation [87,90].

In particular, when performing structure-aided NOE auto-assignment and structure modeling, the quality of the resultant structures strongly depends on the completeness and correctness of the assigned chemical shifts [86–89]. Notably, a number of spurious NOEs such as noise peaks included in the input can lead to wrong NOE assignments, and thereby...
result in incorrect modeled structures. These issues have been intensively discussed in a literature [90]; nevertheless, a sufficient number of correct but weak NOE signals providing long-range distance information is a key feature for accurate and precise structure modeling [90]. In contrast, using the conventional approach of manual NOE assignments and generation of distance restraints, violations in the structure calculation caused by restraint violations may act as indicators to determine whether there are problems in the restraints list. Thus, in the conventional way, when systematic violations arise from the presence of incorrect restraints, the structure calculation and revision of restraints have to be repeated exhaustively. Both conventional and automated approaches can lead to endless and repetitive rounds of calculations, and thus a set of criteria is needed to ensure such a cycle is avoided. Previously, the national project Protein 3000 mainly performed by RIKEN solved a large number of structures by solution NMR using the CYANA and Kujira systems (over 1080 structures) [15,17]. This project revealed a number of benchmarks to achieve when solving structures by NMR: 1) the remaining number of unassigned NOE peaks should be $\leq$5%, which should not be localized to a certain region of the calculated structure; 2) high completeness and accuracy of the assigned chemical shifts of $^1$H, $^{13}$C and $^{15}$N signals ($\geq$90% of completeness); and 3) a sufficient number of NOE signals in the NOESY spectra ($\geq$10–20 NOEs per residue), with a wide intensity range and sufficiently high signal-to-noise ratio and resolution. Using the Kujira and MagRO system, the analysts can easily and visually inspect the accuracy of the assigned chemical shifts and the NOE signal that were automatically assigned by CYANA. In the process, NOE signals showing abnormal line-shapes and/or spurious NOE peaks can be eliminated. The aforementioned strategy should be suitable for structural studies of small proteins that give good-quality NMR spectra, which has been quantitatively evaluated by Kobayashi et al. [15,17].

Furthermore, a systematic violation found in the structure calculations when there is sufficient NOE network anchoring may aid in identifying inappropriate angle restraints.

A general way to model structures using distance and angle restraints is by simulated annealing, as described above [86,90]. The protocol of the calculation has been standardized. A completely unfolded state of the target protein, namely at an extremely high temperature (e.g. 10,000 K), is virtually generated as an initial model, and the temperature is gradually cooled to around 0 K to minimize the total potential energy (or target functions) and to satisfy the input NMR restraints.

It is well known that because the modeled NMR structures calculated by simulated annealing tend to be trapped into local minima of the energy (or target functions) and to satisfy the input NMR restraints. The interaction of the dipolar coupling between two spin-1/2 nuclei depends on the angle and distance between the bond vector and their gyromagnetic ratio. In contrast to solid-state NMR, the Hamiltonian derived from the dipolar coupling interaction between two dipolar-coupled nuclei is not observed because it averages to zero because of isotropic tumbling of the protein in solution. When the protein is dissolved together with a medium that orients under the influence of the magnetic field, e.g., a liquid crystalline formed by lipid discs such as bicelles [156], or filamentous phage such as Pf1 [157,158], the aligned medium will provide some extent of anisotropy to the proteins owing to slight restriction of the isotropic molecular tumbling due to repetitive

7. Validation of Precision and Accuracy of the Determined Protein Structure

Validation of the modeled structures to assess the precision and accuracy is an important final step in the structure determination process by NMR. There is no guarantee that the protein structure determined by solution NMR with high precision matches the same protein structure determined by a different method [149,150]. To finalize the structural study by NMR, the analyst should carefully verify and assess both the precision and accuracy of the determined structure based on particular criteria.

As mentioned above, the convergence (precision) of the modeled structures can be generally assessed through the RMSD values of the 10–30 lowest minimal potential energy structures [151].

The accuracy of the modeled structures can be assessed through geometric parameters derived from the coordinates of the lowest energy structures [149,150,152]. The most widely used analysis examines bond angles, chirality and the side chain rotamer states, as well as the backbone conformation by creation of a Ramachandran plot. The results of the Ramachandran plot analysis are classified into four groups: 1) most favored, 2) additionally allowed, 3) generously allowed and 4) disallowed. Ideally the number of residues found in the generously allowed and disallowed regions of the Ramachandran plot is zero, or at least less than 10%, indicating that there are only a few geometric issues among the atoms and bonds in the structural ensemble [150]. This kind of analysis can be performed by PROCHECK-NMR [153,154] or MolProbity [155], and is a requirement when depositing the molecular coordinates into the PDB.

The orientation of bond vectors in a protein, such as $^1$H$^N$-$^{15}$N and $^1$H$^P$-$^{13}$C, relative to the principal axis of the molecular alignment tensor of the structure can be investigated by measuring residual dipolar coupling (RDC) between the dipolar-coupled nuclei using uniformly $^{15}$N- and/or $^{13}$C-labeled protein in a weakly aligned medium [152]. The information about the bond vector orientations in a protein can be used to validate the accuracy of the determined structure [152]. The interaction of the dipolar coupling between two spin-$1/2$ nuclei depends on the angle and distance between the bond vector and their gyromagnetic ratio. In contrast to solid-state NMR, the Hamiltonian derived from the dipolar coupling interaction between two dipolar-coupled nuclei is not observed because it averages to zero because of isotropic tumbling of the protein in solution. When the protein is dissolved together with a medium that orients under the influence of the magnetic field, e.g., a liquid crystalline formed by lipid discs such as bicelles [156], or filamentous phage such as Pf1 [157,158], the aligned medium will provide some extent of anisotropy to the proteins owing to slight restriction of the isotropic molecular tumbling due to repetitive
collisions between the protein and the aligned media [159]. As a result, a small residual value of the dipolar coupling between nuclei can be observed, and the size of the RDC can be tuned by the extent of alignment of the protein.

From the Cartesian coordinates of the aligned protein, it is theoretically possible to predict the alignment tensor of the weakly aligned protein using programs such as PALES [160] or REDCAT [161] by simulation of the predicted dipolar coupling between the bond vector and the principal axes of the estimated alignment tensor of the protein.

RDC values can also be calculated from the determined structure. The back-calculated RDC values from the structure (the RDC data have not been used as restraints in the structure calculation) should show a high correlation (correlation coefficient \(= 0.8-1.0\)) with the measured RDCs, confirming that the structure has been correctly determined [162]. Therefore, we would encourage this approach as a means to validate the protein structure, if the target protein is stable in alignment media and a sufficient number of RDCs are measured.

As aforementioned, if the analyst wants to directly apply the RDC values as restraints for structure calculation [159], it is necessary that the alignment tensor of the protein be determined precisely and unambiguously in advance. In a conventional manner, precise and unambiguous determination of the alignment tensor of the protein based on RDCs is impossible without information about the tertiary structure of the target protein. Therefore, conventionally, restraints for bond vector orientations using RDCs are generally used during the final refinement stage in NMR structure calculations [159]. Recently, however, methods for estimating the alignment tensor based solely on a histogram distribution of the experimental RDCs (without information about the tertiary structure) have been developed [163].

There is always a certain possibility that the determined structure with a high convergence is (locally or wholly) incorrect because intermolecular distance restraints based on NOE and dihedral angle restraints fundamentally provide local (short-range, \(<6\) Å) information, which means that those restraints cannot unambiguously restrict the relative orientation between each secondary structure or sub-domains of a protein [159]. In particular, in the cases of a multidomain protein or a protein–peptide complex, accurate determination of the relative orientation between each molecule is difficult when using such short-range restraints, even if the tertiary structure of each component has been determined precisely and accurately [164–167]. When the NMR structures have been refined using RDC data as restraints, as a matter of course, the RDC data may no longer be applied to validation.

8. Concluding Remarks

Solution NMR spectroscopy is a popular method to determine the tertiary structure of proteins. Surprisingly, however, there are only a handful of articles that describe the workflow of protein NMR structure determination [71,168–170]. We anticipate that this mini-review should help scientists who are interested in protein solution structure determination by NMR.

The rapid and steady progress in NMR hardware and software, e.g., CYANA upgrades and automation of NMR signal assignment programs, has continued unabated [171]. Furthermore, recently, integrated NMR software platforms offering systematic and semi-automatic biomolecular structure determination, e.g. PONDEROSA-C/S [16,18,126] and UNIO [19,87,88,95–97], have been developed. Moreover, global fold determination of desired proteins has now become easier owing to the development of programs, e.g. CS-ROSETTA [172,173], which can determine protein structures using only chemical shifts and RDC data as restraints.

Furthermore, advanced techniques for protein sample preparation (e.g., SAIL technology [42,174–177], methyl group-selective \(^1\)H,\(^{13}\)C-labeling with deuterium of the other non-labile protoins [178–183]) combined with elaborate NMR pulse schemes (e.g., rapid NMR data collection such as the Band Selective Optimized-Flip-Angle Short-Transient (SOFAST), Band-selective Excitation Short-Transient (BEST) [184,185] and their easy set up/use scripts (http://www.ibs.fr/research/scientific-output/software/pulse-sequence-tools/), ASAP or ALFOSFAST [186] methods, the 3D or 4D methyl-methyl NOESY based on high-resolution and diagonal-free HMOC-NOESY-HMQC pulse schemes [181,187–189], and the dual- or parallel-FID acquisition approaches [190–192], and non-uniform data sampling (NUS) applying NMR data collection of the indirect dimension and quantitative reconstitution of NMR spectra from sparsely sampled data [193–195], should facilitate the study of challenging proteins by NMR (e.g., membrane proteins, enzymes such as kinase/phosphatase, and supramolecular complexes). Therefore, solution NMR spectroscopy is expected to develop even further as a tool for determining the tertiary structure of macromolecules and large molecular weight protein complexes (ca. >50 kDa) [196–200].

In addition, solution NMR spectroscopy is a powerful tool for intermolecular interaction studies and screening of druggable seed compounds with atomic level resolution, because it can detect and determine a binding site, even if the interaction is extremely weak. By performing magnetization relaxation dispersion NMR experiments, tertiary structures of minor, excited or invisible protein conformations can be determined when this conformation is populated by \(<1\%\) [201,202]. Therefore, it is expected that methodological advances of solution NMR spectroscopy, which can determine the tertiary structure of a protein in the bound state accurately and precisely in a weak interaction, will become a unique and beneficial structural biology tool by maximizing the features of solution NMR spectroscopy. Experimentally determined intermolecular distances and secondary structure information predicted by TALOS and its related programs [84,85], CamShift [120] and SHIFTX2 [121] can greatly assist accurate homology based modeling or docking simulations to obtain structural models. The most successful examples using these methods would be HADDOCK [203,204] and CS-ROSETTA [172,173], whose feasibility has been generally accepted to yield structural insights that give rational explanations describing the biological functions of biomolecular complexes and models. The integration of the above methods and/or additional structural data such as X-ray scattering and information about intermolecular contacts derived from pull-down and yeast-two-hybrid assays as well as point mutation data, the so-called hybrid/integrative methods, is a new approach to study large biomacromolecular systems [205].

A number of software and web-based resources for NMR data analysis as described in this article are available and user-friendly, and should help to open avenues for non-specialist and life scientists interested in studying the structure of proteins.

Acknowledgements

We are grateful to Dr. Kyoko Furuita, Dr. Shoko Shinya, Prof. Young-Ho Lee (Osaka University, Japan), Dr. Yoshikazu Hattori (Tokushima Bunri University, Japan) and Prof. Chojiro Kojima (Yokohama National University, Japan) for many useful discussions and suggestions. This work was supported in part by the Platform Project for Supporting Drug Discovery and Life Science Research, and AMED-CREST from MEXT and AMED, Japan.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.csbj.2017.04.001.

References

[1] Ilari A, Savino C. Protein structure determination by X-ray crystallography. Methods Mol Biol 2008;452:63–87.
[2] Wüthrich K. The way to NMR structures of proteins. Nat Struct Biol 2001;8:923–5.
[3] Zhou ZH. Atomic resolution cryo electron microscopy of macromolecular complexes. Adv Protein Chem Struct Biol 2011;82:1–35.
Liu Z, Gong Z, Dong X, Tang C. Transient protein-protein interactions visualized by Rasia R, Brutscher B, Plevin M. Selective isotopic unlabeling of proteins using metabolic approaches. J Biomol NMR 2007;37:225–32.

Takai K, Sasaki T, Endo Y. Practical cell-free protein synthesis system using purified wheat embryo. Nat Protoc 2010;5:227–38.

Yokoyama J, Matsuura T, Shigai K. A mathematical model for predicting stable isoform labeled proteins by the E. coli cell-free system. J Biomol NMR 2010;48:193–201.

Hiroi L, Ohtsuki T, Fujitawa T, Mitsui T, Yokogawa T, Okuni T, et al. Unnatural base pair for incorporating amino acid analogs into proteins. Nat Biotechnol 2011;29:177–82.

Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Os, Gunter P. Optimal iso- tope labeling for NMR protein structure determinations. Nature 2006;440:52–7.

Abe R, Shiraga K, Eislie S, Takagi H, Holaska S. Incorporation of fluorescent non-natural amino acids into N-terminal tag of proteins in cell-free translation and its dependance on position and neighboring codons. J Biosci Bioeng 2010;110:32–8.

Loscha RV, Herli AJ, QI R, Huber T, Ozawa K, Gunter M. Single-site labeling of proteins with unnatural amino acids. Angew Chem Int Ed 2012;51:2243–6.

Gu X, Liu CQ, QI Q, Tang H, Zhu J, Wang B, et al. Stable protein synthesis by broad inhibition of PLE enzymes for selective 13N-labeling and production of deuterium labeled proteins in H2O. J Biomol NMR 2011;50:35–42.

Shimizu Y, Karuma Y, Kanamori T, Ueda T. The PURE system for protein production. Methods Mol Biol 2010;567:175–94.

Deniaux A, Liguori I, Blesnie C, Lenormand JL, Peyba-Peyroula E. Crystallization of the membrane protein hVDAC1 produced in cell-free system. Biochim Biophys Acta 2010;1798:1540–6.

Reckel S, Sobhanifar S, Furst D, Loh F, Shirakova VA, Ditsch V, et al. Strategies for the cell-free expression of membrane proteins. Methods Mol Biol 2010;607:187–212.

Habestock S, Roos C, Hoevels Y, Wüthrich K. NMR protein structure determination in cell-free expression systems. Protein Expr Purif 2012;82:308–16.

Wang X, Wang J, Ge B. Evaluation of cell-free expression system for the production of soluble and functional human GPCR N-formyl peptide receptor. Protein Lett 2013;20:1272–9.

Dutta A, Saxena K, Schwabie H, Klein-Seetharaman J. Isopeptide labeling in mammalian cell systems. Methods Mol Biol 2012;831:195–59.

Saxtry M, Bewley CA, Kowg MD. Potential metabolic expression of isotopically labeled proteins for NMR spectroscopy. Adv Exp Med Biol 2012;992:179–211.

Kikuchi J, Shinozaki K, Hirayama T. Stable isotope labeling of Arabidopsis thaliana for an NMR-based metabolomics approach. Plant Cell Physiol 2004;45:1099–104.

Brezde-Efardjan T, Herrmann T, Horst R, Wüthrich K. Automated protein NMR structure determination in crude cell-extract. J Biomol NMR 2006;34:3–11.

Chikayama E, Suto M, Nishihara T, Shinozaki K, Hirayama T, Kikuchi J. Systematic NMR analysis of stable isotope labeled metabolite mixtures in plant and animal cells using 13C NMR measurements. J Biomol NMR 2008;39:385–95.

Latham LP, Kay LE. Is bench a good proxy for a crowded cell-like environment? A comparative NMR study of calmodulin-side chain dynamics in buffer and E. coli lysate. PLoS One 2012;7:e48226.

Tomita S, Beda S, Tani M, Takayama N, Asano K, Kikuchi J, et al. A survey of metabolic changes in potato leaves by NMR-based metabolic profiling in relation to resistance to late blight disease under field conditions. Magn Reson Chem 2013;55:120–7.

Zheng Z, Sun H, Hu C, Li G, Liu X, Chen P, et al. Using “on-off” (19)F NMR magnetic resonance imaging signals to sense tyrosine kinase phosphatase activity in vitro and in cell lysates. Anal Chem 2016;88:3368–33.

Sugiki T, Yoshuara C, Kohfuku Y, Ueda T, Shimada I, Takahashi H. High-throughput screening of optimal solution conditions for structural biological studies by fluorescence correlation spectroscopy. Protein Sci 2009;18:1115–20.

Horst R, Wüthrich K. Micro-scale NMR experiments for monitoring the optimization of membrane protein solutions for structural biology. Bio Protoc 2015; 5: e1153.

Kozak S, Lercher J, Karahan MN, Meijers R, Carlnagmo T, Boivin S. Optimization of protein samples for NMR using thermal shift assays. J Biomol NMR 2016;64:281–9.

Pedriani B, Serrano P, Mohanty B, Geraldt M, Wüthrich K. NMR-profiles of protein solutions. Biopolymers 2013;99:285–31.

Flynn P, Mattiodo D, Hill H, Wand A. Optimal use of cysogenic probe technology in NMR studies of proteins. J Am Chem Soc 2000;122:4823–4.

Lee D, Hilly C, Wider C, Wüthrich K. Effective rotational correlation times of protein samples from NMR relaxation measurements. J Mag Reson 2008;197:72–2.

Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by 15N magnetic resonance spectroscopy. J Am Chem Soc 2003;125:7174–7.

Sugiki T, Ishiwa O, Miyazawa-Onami M, Shimada I, Takahashi H. Isotope labeling of heterologous proteins in the yeast Pichia pastoris and Klyuyveromyces lactis. Metab. Mol Biol Biotechnol 2012;8:119–36.

Kelly AE, Ou HD, Withers R, Dötsch V. Low-conductivity buffers for high-sensitivity NMR measurements. J Am Chem Soc 2002;124:12031–9.
[69] Driscoll PC, Clere GM, Marion D, Wingfield PT and Gronenborn AM (1990) Complete resonance assignment for the polypeptide backbone of interleukin 1 beta using three-dimensional heteronuclear NMR spectroscopy. Biochemistry 29: 3524-3530.

[70] Clere GM, Bax A, Driscoll PC, Wingfield PT, Gronenborn AM. Assignment of the side-chain 1H and 13C resonances of interleukin-1 beta using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy. Biochemistry 1990;29:8172-84.

[71] Kanelis V, Forman-Kay JD, Kay LE. Multidimensional NMR methods for protein structure determination. IUBMB Life 2001;52:291–302.

[72] Lee W, Cornilescu G, Dashti H, Eghbalnia HR, Tonelli M, Westler WM, et al. Automated protein structure determination using nuclear magnetic resonance. Proc Natl Acad Sci USA 1997;94:12366-12371.

[73] Vögeli B, Olsson S, Güntert P, Riek E. The exact NOE as an alternative in ensemble protein-protein complexes. J Biomol NMR 2010;46:271-80.

[74] Tikole S, Jaravine V, Orekhov V, Güntert P. Effects of NMR spectral resolution on protein structure calculation. PloS One 2013;8:e65657.

[75] Vögeli B, Olsson S, Güntert P. Structure determination of protein-protein complexes. J Biomol NMR 2010;46:271-80.

[76] Vögeli B, Olsson S, Güntert P. The exact NOE as an alternative in ensemble protein-protein complexes. J Biomol NMR 2010;46:271-80.

[77] Vögeli B, Olsson S, Güntert P. The exact NOE as an alternative in ensemble protein-protein complexes. J Biomol NMR 2010;46:271-80.

[78] Würz JM, Güntert P. Peak picking multidimensional NMR spectra with the contour integration based algorithm CYPICK. J Biomol NMR 2017;67:63–76.

[79] Würz JM, Güntert P. Peak picking multidimensional NMR spectra with the contour integration based algorithm CYPICK. J Biomol NMR 2017;67:63–76.

[80] Englander SW, Mayne L. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. Annu Rev Biophys Biomol Struct 1992;21:243–65.

[81] Würz JM, van Zijl PC, van Sint Annaland B, Wüthrich K. Automated quantification of water-amine proton exchange rates using the phase-modulated clean chemical exchange (CLEANEX-M) approach with a Fast-HSQC (FHSQC) detection scheme. J Biomol NMR 1998;11:221–6.

[82] Saio T, Yokochi M, Kumeta H, Inagaki F. PCS-based structure determination of very large biological macromolecules.Curr Opin Struct Biol 2003;13:570–80.

[83] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional NMR processing system for isolation of nuclear resonance peaks. J Magn Reson 1995;128:277–93.

[84] Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, et al. Recommendations for presentation of NMR structures of proteins and nucleic acids. IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of proteins and nucleic acids. J Mol Biol 1995;254:465–72.

[85] Vögeli B, Olsson S, Güntert P. Automated NOE assignment. J Struct Funct Genomics 2003;4:179-89.

[86] Güntert P. Automated NMR structure calculation with CYANA. Methods Mol Biol 2004;278:353-78.

[87] Kortij R, Billeter M, Engeli M, Güntert P, Wüthrich K. Automated picking of backbone nuclei in unfolded proteins. J Magn Reson 1998;135:288-93.

[88] Würz JM, Güntert P. Peak picking multidimensional NMR spectra with the contour integration based algorithm CYPICK. J Biomol NMR 2017;67:63–76.

[89] Englander SW, Mayne L. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. Annu Rev Biophys Biomol Struct 1992;21:243–65.

[90] Vögeli B, Olsson S, Güntert P. The exact NOE as an alternative in ensemble protein-protein complexes. J Biomol NMR 2010;46:271-80.

[91] Kortij R, Billeter M, Engeli M, Güntert P, Wüthrich K. Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY. J Magn Reson 2011;215:288-97.

[92] Kortij R, Billeter M, Engeli M, Güntert P, Wüthrich K. Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY. J Magn Reson 2011;215:288-97.

[93] Vögeli B, Olsson S, Güntert P. The exact NOE as an alternative in ensemble protein-protein complexes. J Biomol NMR 2010;46:271-80.

[94] Vögeli B, Olsson S, Güntert P. The exact NOE as an alternative in ensemble protein-protein complexes. J Biomol NMR 2010;46:271-80.

[95] Würz JM, Güntert P. Peak picking multidimensional NMR spectra with the contour integration based algorithm CYPICK. J Biomol NMR 2017;67:63–76.

[96] Englander SW, Mayne L. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. Annu Rev Biophys Biomol Struct 1992;21:243–65.

[97] Würz JM, van Zijl PC, van Sint Annaland B, Wüthrich K. Automated quantification of water-amine proton exchange rates using the phase-modulated clean chemical exchange (CLEANEX-M) approach with a Fast-HSQC (FHSQC) detection scheme. J Biomol NMR 1998;11:221–6.

[98] Wang YX, Jacob J, Cordier F, Wingfield P, Stallf SJ, Lee-Huang S, et al. Measurement of 15N chemical shifts using artificial neural networks. J Biomol NMR 1995;6:135–40.

[99] Markley BA, Blevins RA. NMR View: a computer program for the visualization and analysis of NMR data. J Biomol NMR 1994;4:603-14.

[100] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[101] Wang YX, Jacob J, Cordier F, Wingfield P, Stallf SJ, Lee-Huang S, et al. Measurement of 15N chemical shifts using artificial neural networks. J Biomol NMR 1995;6:135–40.

[102] Vögeli B, Orts J, Strotz D, Chi C, Minges M, Wälti MA, et al. Towards a true protein chemical shift database: the reevaluation of the solution state 13C- and 15N chemical shifts for proteins with sparse NOE distance restraints. J Biomol NMR 2015;61:55-64.

[103] Vögeli B, Olsson S, Güntert P, Riek E. The exact NOE as an alternative in ensemble structure determination. Bioparts 2016;110:113–26.

[104] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[105] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[106] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[107] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[108] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[109] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[110] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[111] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[112] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[113] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.

[114] Hu H, Krishnamurthy K. Revisiting the initial rate approximation in kinetic NOE measurements. J Magn Reson 2006;182:173-7.
Nabuurs CD, Clore GM. The VMD-XPLOR visualization package for NMR structures. J Biomol NMR 2014;59:135–45.

Kuszewski CD, Schwiderski CD, Garrett DS, Byrd RA, Tjandra N, Clore GM. Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear Overhauser enhancement spectra and chemical shift assignments. J Am Chem Soc 2004;126:6258–73.

Schwiderski CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. J Magn Reson 2003;160:65–73.

Brunger A, Adams P, Clore G, DeLano W, Gris P, Grosse-Kunstleve R, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Cryst Soc F 1998;54:905–21.

Salomon-Ferrer R, Case D, Walker R. An overview of the Amber biomolecular simulation package, Wiley Interdiscip Rev Compu Mol Sci 2013;3:198–210.

Robertson MJ, Tirado-Rives J, Jorgensen WL. Improved peptide and protein torsion-angle energetics with the OPLS force field. J Chem Theory Comput 2011;7:4399–509.

Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 2008;29:1853–69.

Kelley L, Gardner S, Sutcliffe M. An automated approach for defining core domains and domains in an ensemble of NMR-derived protein structures. Prot Eng 1997;10:737–41.

Snyder DA, Gruhn J, Huang YJ, Tejero R, Montelione GT. The expanded findcore method for identification of a core atom set for assessment of protein structure prediction. Proteins 2014;82:219–30.

Kirkner DK, Güntert P. Objective identification of residue ranges for the superposition of protein structures. BMC Bioinformatics 2011;12:170.

Kobayashi N. A robust method for quantitative identification of ordered cores in an ensemble of biomolecular structures by non-linear multi-dimensional scaling using inter-atomic distance variance matrix. J Biomol NMR 2014;58:413–25.

Petterson EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera — a visualization system for exploratory r y and analysis. J Comput Chem 2004;25:1605–12.

McNicholas S, Petterson E, Wilson KS, Noble ME. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystalllography D 2011;67:386–94.

Schwiderski CD, Clore GM. The VMD-XPLOR visualization package for NMR structure refinement. J Magn Reson 2001;149:239–44.

Kosari R, Billette K, Ballesteros C. LaMOL for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–5.

Nabuurs S, Spronk C, Friend G, Vuister G. Concepts and tools for NMR restraint analysis and validation. Concept Magn Reson Part A 2004;22A:90–105.

Spronk C, Nabuurs S, Krieger E, Friend G, Vuister G. Validation of protein structures derived by NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 2004;45:315–37.

Huang X, Myo F, Powers R. Evaluation of the utility of NMR structures determined from minimal NOE-based restraints for structure-based drug design, using MMP-1 as an example. Biochemistry 2000;39:13165–75.

Rosato A, Tejero R, Montelione G. Quality assessment of utility of NMR structures. Curr Opin Struct Biol 2013;23:715–24.

Laskowski R, MacArthur M, Moss D, Thornton J. PROCHECK—a program to check the stereochemical quality of protein structures. J Appl Crystrallography 1993;26:283–91.

Laskowski R, Rullmann J, MacArthur M, Kaptein R, Thornton J. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;6:477–86.

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom stereochemical analysis and validation. Concept Magn Reson Part A 2004;22A:90–105.

Bax A, Tjandra N. High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. J Biol Chem 1997;272:1089–92.

Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by flourescent phase plates yields dipolar coupling interactions. Nat Struct Biol 1998;5:1065–74.

Zweckstetter M, Bax A. Characterization of molecular alignment in aqueous suspensions of PTD biophotographs. J Biol Chem 2001;276:305–77.

Lipskii I, Tjandra N. Residual dipolar coupling in NMR structure analysis. Annu Rev Biochem 2004;73:387–412.

Zweckstetter M, NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 2008;3:679–90.

Schulze-Sünninghausen D, Becker J, Luy B. Rapid heteronuclear single quantum correlation NMR spectra at normal abundance. J Am Chem Soc 2014;136:12442–5.

Laskowski R, Ibraghimov I, Orekhov V. High-resolution four-dimensional H-1–C-13 NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional projection for the determination of large proteins. Biochemistry 2009;48:477–86.

Han D, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom alignments and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35:W175–83.

Bax A, Tjandra N. High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. J Biol Chem 1997;272:1089–92.

Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by flourescent phase plates yields dipolar coupling interactions. Nat Struct Biol 1998;5:1065–74.

Zweckstetter M, Bax A. Characterization of molecular alignment in aqueous suspensions of PTD biophotographs. J Biol Chem 2001;276:305–77.

Lipskii I, Tjandra N. Residual dipolar coupling in NMR structure analysis. Annu Rev Biochem 2004;73:387–412.

Zweckstetter M, NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 2008;3:679–90.

Schulze-Sünninghausen D, Becker J, Luy B. Rapid heteronuclear single quantum correlation NMR spectra at normal abundance. J Am Chem Soc 2014;136:12442–5.

Laskowski R, Ibraghimov I, Orekhov V. High-resolution four-dimensional H-1–C-13 NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional projection for the determination of large proteins. Biochemistry 2009;48:477–86.

Bax A, Tjandra N. High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. J Biol Chem 1997;272:1089–92.

Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by flourescent phase plates yields dipolar coupling interactions. Nat Struct Biol 1998;5:1065–74.

Zweckstetter M, Bax A. Characterization of molecular alignment in aqueous suspensions of PTD biophotographs. J Biol Chem 2001;276:305–77.

Lipskii I, Tjandra N. Residual dipolar coupling in NMR structure analysis. Annu Rev Biochem 2004;73:387–412.

Zweckstetter M, NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 2008;3:679–90.

Schulze-Sünninghausen D, Becker J, Luy B. Rapid heteronuclear single quantum correlation NMR spectra at normal abundance. J Am Chem Soc 2014;136:12442–5.

Laskowski R, Ibraghimov I, Orekhov V. High-resolution four-dimensional H-1–C-13 NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional projection for the determination of large proteins. Biochemistry 2009;48:477–86.

Bax A, Tjandra N. High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. J Biol Chem 1997;272:1089–92.

Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by flourescent phase plates yields dipolar coupling interactions. Nat Struct Biol 1998;5:1065–74.

Zweckstetter M, Bax A. Characterization of molecular alignment in aqueous suspensions of PTD biophotographs. J Biol Chem 2001;276:305–77.

Lipskii I, Tjandra N. Residual dipolar coupling in NMR structure analysis. Annu Rev Biochem 2004;73:387–412.

Zweckstetter M, NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 2008;3:679–90.

Schulze-Sünninghausen D, Becker J, Luy B. Rapid heteronuclear single quantum correlation NMR spectra at normal abundance. J Am Chem Soc 2014;136:12442–5.
