the FEBS Journal

REVIEW ARTICLE

Macromolecular NMR spectroscopy for the non-spectroscopist

Ann H. Kwan[1,*], Mehdi Mobli[2,*], Paul R. Gooley[3], Glenn F. King[2] and Joel P. Mackay[1]

1 School of Molecular Bioscience, University of Sydney, New South Wales, Australia
2 Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland, Australia
3 Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria, Australia

Keywords
HSQC; nuclear magnetic resonance (NMR) spectroscopy; protein folding; protein NMR spectroscopy; protein stability; protein structure determination; TROSY

Correspondence
J. P. Mackay or G. F. King, School of Molecular Bioscience, University of Sydney, Sydney, NSW 2006 Australia; Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD 4072, Australia
Fax: +61 2 9351 4726; +61 7 3346 2101
Tel: +61 2 9351 3906; +61 7 3346 2025
E-mail: joel.mackay@sydney.edu.au; glenn.king@imb.uq.edu.au

*These authors contributed equally to this work

(Received 20 July 2010, revised 7 November 2010, accepted 5 January 2011)

doi:10.1111/j.1742-4658.2011.08004.x

NMR spectroscopy is a powerful tool for studying the structure, function and dynamics of biological macromolecules. However, non-spectroscopists often find NMR theory daunting and data interpretation nontrivial. As the first of two back-to-back reviews on NMR spectroscopy aimed at non-spectroscopists, the present review first provides an introduction to the basics of macromolecular NMR spectroscopy, including a discussion of typical sample requirements and what information can be obtained from simple NMR experiments. We then review the use of NMR spectroscopy for determining the 3D structures of macromolecules and examine how to judge the quality of NMR-derived structures.

Introduction

NMR spectroscopy is a powerful tool for the analysis of macromolecular structure and function. Approximately 8300 NMR-derived protein structures have now been deposited in the Protein Data Bank (PDB). Moreover, a number of methodological and instrumental advances over the last 20 years or so have dramatically increased the breadth of biological problems to which NMR spectroscopy can be applied. Although the theory underlying the phenomenon of NMR spectroscopy is daunting (even to many NMR spectroscopists!), a background in quantum mechanics is not required to gain a good appreciation of what information is contained in an NMR spectrum, as well as the strengths, limitations and requirements of the technique.

In this review, we provide an introduction to the principles of macromolecular NMR spectroscopy, including basic interpretation of commonly encountered NMR spectra. We then outline the process by

Abbreviations
PDB, Protein Data Bank; RDC, residual dipolar coupling; RMD, restrained molecular dynamics; TROSY, transverse relaxation optimized spectroscopy.

which NMR is used to determine the 3D structure of a protein or nucleic acid in solution. Finally, we focus on how to assess the quality of a published structure, as well as the sort of information that the structure can provide. Biomolecular NMR spectroscopy is not, however, restricted to macromolecular structure determination, and the breadth of biological questions that can be addressed using NMR is probably unparallelled by any other form of spectroscopy. In the accompanying review [1], we introduce the reader to some of the more common applications of NMR for understanding macromolecular function.

Throughout these reviews, we have attempted to highlight the strengths and weaknesses of NMR spectroscopy and, where appropriate, make reference to complementary techniques. We hope that these reviews can help to alert researchers in the life sciences to the power and relatively straightforward nature of NMR approaches and allow them to better evaluate NMR data reported in the literature.

NMR for everyone

The NMR phenomenon: a potted summary

Similar to all forms of spectroscopy, NMR spectra can be considered to arise from transitions made by atomic nuclei between different energy states (indeed, this is an oversimplification, although this need not concern us here; for more details, see Keeler [2]). For reasons that we will not go into, the nuclei of many isotopes such as ^{1}H, ^{13}C, ^{15}N and ^{31}P carry magnetic dipoles. These dipoles take up different orientations in a magnetic field, such as the magnet of an NMR spectrometer, and each orientation has a different energy. Transitions between states with certain energies are permitted according to the postulates of quantum mechanics and, when we apply pulses of electromagnetic radiation at frequencies that precisely match these energy gaps, we are able to observe transitions that give rise to NMR signals. Nuclei in different chemical environments (e.g. the different ^{1}H nuclei in a protein) will resonate at different frequencies and a plot of intensity against resonance frequency is known as a 1D NMR spectrum. Resonance frequencies are typically reported as 'chemical shifts' in units of p.p.m., which corrects for the fact that the raw frequencies (usually in units of MHz) scale with the size of the NMR magnet.

One of the key features that differentiates NMR from most other forms of spectroscopy is that the excited states are relatively long lived, with lifetimes in the millisecond–second range (in contrast to the nanosecond timescales that define fluorescence or infrared spectroscopy). Consequently, we can manipulate the excited state to pass excitation from one nucleus to another and, indeed, multiple transfer steps are common in a single experiment. Because we can measure the frequencies of each of the nuclei through which excitation (magnetization) is passed, we can obtain signals that correlate (link) the frequencies of two, three or more nuclei. In such correlation spectra, each transfer can be visualized as an independent nuclear frequency dimension (axis) and signals occurring at the intersection of two or more frequencies indicate a correlation between the corresponding nuclei. The resulting multidimensional spectra allow us to determine unambiguously which signal in a spectrum arises from which atom in the molecule. This process of frequency assignment is an essential step in extracting structural or functional information about the system.

For a detailed account of NMR theory, we recommend the books by Keeler and Levitt [2,3], as well as the monograph by Cavanagh *et al.* [4], which is focused entirely on protein NMR spectroscopy.

Your first NMR spectra

Two of the most useful and sensitive NMR spectra are the 1D ^{1}H-NMR spectrum (Fig. 1A), which simply shows signals for each of the hydrogen atoms (referred to as 'protons' in the NMR world) in a biomolecule, and the 2D ^{15}N-HSQC (heteronuclear single-quantum coherence) spectrum, which shows a signal for each covalently bonded ^{1}H-^{15}N group [5] (Fig. 1B). Each signal in this latter spectrum has an intensity and two chemical shifts (one for the ^{1}H and another for the ^{15}N nucleus) and the spectrum is plotted 'looking from above', much like a topographic map. For a well-behaved protein, the ^{15}N-HSQC spectrum will contain one peak for each backbone amide proton (i.e. one for each peptide bond, except those preceding prolines), a peak for each indole NH of tryptophan residues, and pairs of peaks for the sidechain amide groups of each Asn and Gln residue (for these amide groups, each ^{15}N nucleus has two attached protons). Under favourable circumstances, signals from the guanidino groups of arginine can also be observed. In essence, the ^{15}N-HSQC spectrum should contain one peak for each residue in the protein and, consequently, this spectrum provides an excellent high-resolution 'fingerprint' of the protein. Similarly, a ^{13}C-HSQC spectrum displays a signal for each covalently bonded ^{1}H-^{13}C pair (Fig. 1C). The peaks in this spectrum are not as well resolved as those in a ^{15}N-HSQC spectrum because, unlike ^{15}N shifts, both ^{1}H and ^{13}C chemical shifts are strongly correlated with protein secondary structure and hence with each other.

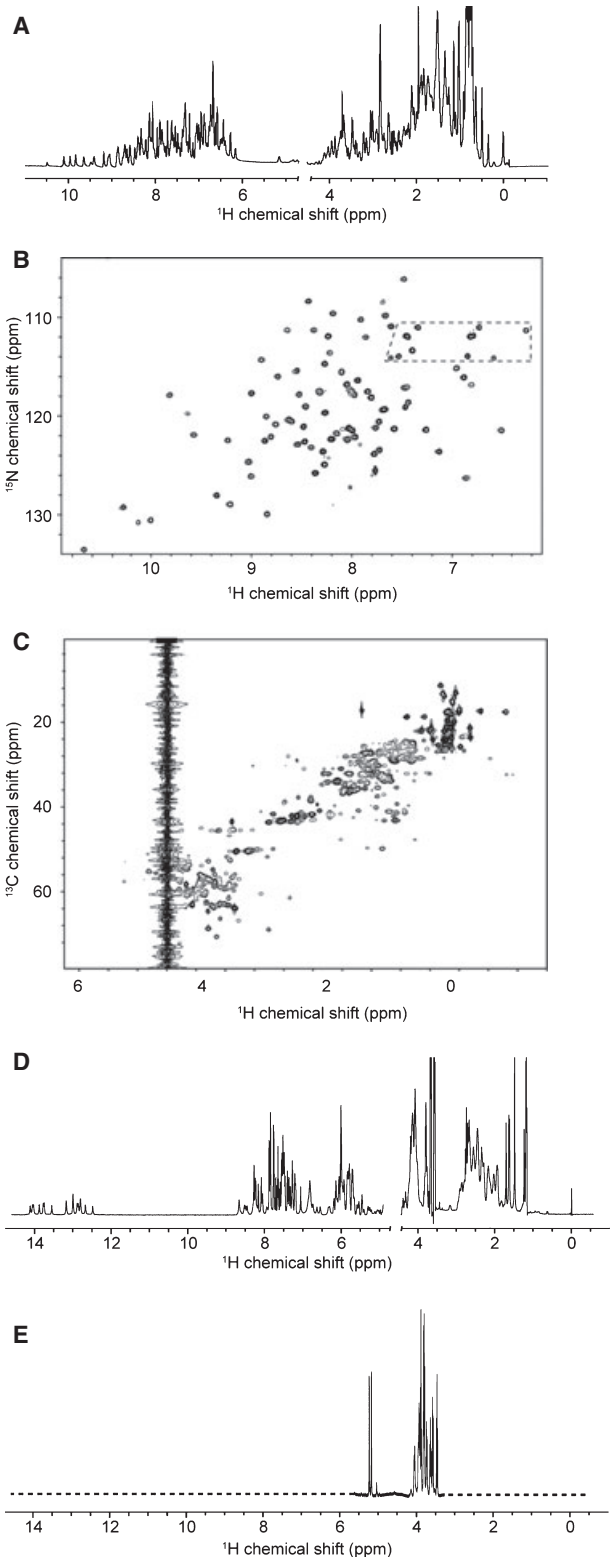

Fig. 1. (A) 1D ^{1}H-NMR spectrum, (B) ^{15}N-HSQC spectrum and (C) ^{13}C-HSQC spectrum of CtBP-THAP, a 10.6 kDa protein. Sidechain amide groups from Asn and Gln residues are indicated by dotted lines. All three spectra were recorded on a 1 mM sample in 20 mM sodium phosphate (pH 6.5) containing 100 mM NaCl and 1 mM dithiothreitol at 298 K on a Bruker 600 MHz spectrometer (Bruker, Karlsruhen, Germany) equipped with a cryoprobe. The spectrum in (A) was recorded over 30 s, whereas the ^{13}C- and ^{15}N-HSQC spectra were recorded over 5 min. (D) 1D ^{1}H-NMR spectrum of a 19 bp (11.7 kDa) double-stranded DNA oligonucleotide. (E) 1D ^{1}H-NMR spectrum of a polysaccharide. Note the poor signal dispersion compared to the protein spectrum.

For comparison, Fig. 1D also shows a 1D ^{1}H-NMR spectrum of a 19 bp, 11.7 kDa double-stranded DNA oligonucleotide. Far fewer signals are observed compared to a protein of the same molecular weight because the nucleotide bases are only sparsely populated with protons. Consequently, it is generally more challenging to carry out detailed NMR-based structural analyses of oligonucleotides compared to proteins. The 1D ^{1}H-NMR spectrum of a polysaccharide is shown in Fig. 1E; the poor dispersion of signals, resulting in severe spectral overlap, combined with difficulties in isotopic labelling, account in part for the dearth of NMR studies of saccharides compared to proteins.

How much sample do I need?

This is one of the first questions asked by potential NMR users. NMR is traditionally known as an information-rich but insensitive form of spectroscopy. Concentrations of approximately 1 mM and sample volumes of approximately 0.5 mL were the typical requirement until relatively recently, restricting NMR to a relatively small fraction of well-behaved, highly soluble molecules. However, hardware advances, in particular the development of higher field magnets and cooled sample detection systems (which reduce electronic noise) [6], have broadened the range of samples that can be studied using NMR methods.

We routinely collect 1D ^{1}H- and 2D ^{15}N-HSQC spectra on 100 μL samples at concentrations of 50 μM; this equates to only 50 μg of a 10 kDa protein. The sample requirements are similar for a ^{13}C-HSQC spectrum. Note also that the sample can be recovered in its entirety subsequent to the recording of data and can be used for other experiments. In comparison, one would typically use approximately 50 μg of a protein (irrespective of molecular weight) to record a far-UV CD spectrum [7] or measure binding events using isothermal titration calorimetry or surface plasmon resonance.

The natural abundances of ^{15}N and ^{13}C isotopes are low (0.4% and 1.1%, respectively) and therefore NMR spectra that measure these nuclei (such as the HSQC spectra mentioned above) are almost exclusively

recorded on recombinant proteins that have been overproduced in a defined minimal medium containing nutrients enriched in these isotopes [e.g. ^{13}C-glucose and $^{15}NH_4Cl$]. Of course, a protein cannot always be produced recombinantly in bacteria, and isotopic labels are not as economically incorporated into other expression systems, although there are exceptions [8]. In this case, it is sometimes possible (but not often feasible) to work at the 'natural abundance' that is provided by nature. The reduction in sensitivity that results in this situation makes recording spectra impractical for all but the most soluble proteins (> 1 mM).

What are the sample requirements?

In general, the sample should be homogeneous (90% purity or greater is preferable). However, NMR work is also routinely carried out on complex mixtures of unknown composition (e.g. in the field of metabolomics) [9]. Although solids can be tolerated in the sample because NMR wavelengths are much longer than typical particle sizes, it is good practice to remove particulates, if only to prevent the nucleation of further aggregation. We note in passing that much biological NMR work has been carried out on suspensions, such as real-time studies of cellular metabolism [10]. It is also worth noting that proteins in the solid state (e.g. microcrystals) have become amenable to detailed NMR studies over recent years; examples are provided by Lesage [11], as well as in the accompanying review [1].

In principle, all buffers are compatible with NMR work. Buffers with many protons will interfere with ^{1}H-NMR spectra, although they will not be a problem when recording spectra (such as a ^{15}N-HSQC) on isotopically labelled samples (because protons not attached to the labelled heteronuclei are 'filtered out'). Minimizing buffer concentrations (approximately 10–20 mM) can be helpful, and deuterated forms of many common buffers are also available. NMR spectra can be recorded at any pH value, with one major caveat. Protons that are chemically labile (such as backbone and sidechain amide protons) can exchange with solvent protons and the rate of this exchange process increases logarithmically at above approximately pH 2.6. Once the exchange becomes sufficiently fast, the signal from a labile proton will merge with that of the solvent and cease to be observable. In practical terms, NMR spectroscopists tend to avoid pH values higher than 7.5 because spectral quality is impaired at higher pH values (Fig. 2). A number of other factors, including the presence of reducing agents, stabilizing agents (such as glycerol) and paramagnetic moieties, also need to be considered.

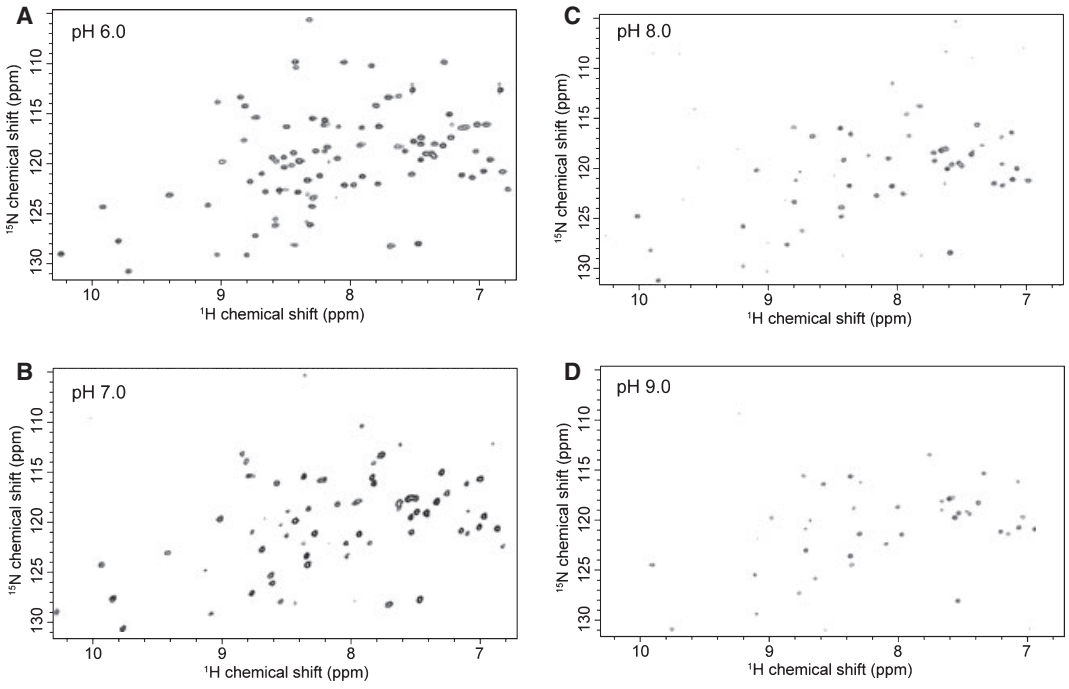

Fig. 2. ^{15}N-HSQC spectra of a 10 kDa polypeptide derived from the zinc-finger protein EKLF, recorded at pH values of (A) 6.0, (B) 7.0, (C) 8.0 and (D) 9.0. Note the decrease in the number of signals from backbone amide protons as the pH is increased.

What information can be deduced from a simple NMR experiment?

Irrespective of whether the aim is to embark on detailed NMR-based structural or functional investigations of a protein, NMR spectroscopy is an excellent (and under-utilized) first-pass quality control method for any sort of biophysical or biochemical programme of research. Armed with a simple 1D ^{1}H-NMR and ^{15}N-HSQC spectrum, there are a number of questions that can be readily answered to provide valuable information for the crystallographer, the enzymologist or the protein engineer. Below, we discuss some common questions that NMR can be used to address.

Is my protein folded?

Figure 3(A, B, C) shows the 1D ^{1}H- and ^{15}N-HSQC spectrum of proteins that are comprised of predominantly α-helix, β-sheet or disordered regions, respectively. The poor signal dispersion displayed by the unfolded protein results from the fact that all amide protons are in similar chemical environments (i.e. exposed to solvent). Spectra of α-helix-rich proteins are also less well dispersed than those from β-sheet-rich proteins as a result of the wider variety of chemical environments found in a β-sheet. Figure 3D shows the spectra for a protein that contains a mixture of well-ordered and completely disordered segments. A count of the number of signals in the disordered or 'random-coil' region of the spectrum (indicated by asterisks) provides a good indication of the fraction of the protein chain that is disordered. This type of simple analysis can provide valuable information for the X-ray crystallographer by alerting them to the presence of disordered regions that might impede crystallization. Assignment of resonances in the ^{15}N-HSQC spectrum (see below) can then provide site-specific information regarding which residues are disordered and which could therefore be targeted for deletion.

Although the spectra of both folded and completely unfolded proteins exhibit sharp lines, proteins that are partially folded often give rise to very poor quality spectra (Fig. 3E). The long-lived excited state in an NMR experiment results in narrow lines with well-defined frequencies (hence the inherently high resolution of the NMR experiment, with linewidths down to approximately 0.1 Hz for small molecules, compared to linewidths of approximately 10^{6} Hz for fluorescence spectra). However, nuclei for which the signal decays more rapidly give rise to broader lines. Interconversion of a protein between different conformations on the μs–ms timescale can cause line broadening of this type. Unexpectedly, such partially folded proteins can often exhibit substantial secondary structure in a far-UV CD spectrum, and a poor quality NMR spectrum can indicate the existence of a so-called molten globule state [12] in which relatively well-formed secondary structural elements are not packed tightly together into a well-defined tertiary structure. Analysis of the ^{15}N-HSQC spectrum will also allow determination of whether the protein is suitable for more detailed NMR-based structural analysis.

Is my protein aggregated?

As noted above, nuclei for which the signal decays more rapidly give rise to broader lines. Slower molecular reorientation also is a major cause of rapid signal decay and therefore broad lines. Self-association will broaden almost all signals, whereas conformational exchange (e.g. between monomer and dimer or bound and free states) will broaden only the signals from the nuclei whose environment is altered by the exchange process (e.g. those at a protein–ligand interface). It can, however, be difficult to distinguish between these two situations from NMR spectra alone and, if presented with an unexpectedly broad spectrum, it is best to examine the aggregation state of the protein further using gel filtration (preferably in conjunction with multi-angle laser light scattering), dynamic light scattering or analytical ultracentrifugation.

Is my protein dynamic?

Counting the signals in the ^{15}N-HSQC spectrum will often reveal dynamic processes. For example, Fig. 3F shows the ^{15}N-HSQC of YPM, a 119 residue (14 kDa) superantigen from *Yersinia pseudotububerculosis* [13]. Although approximately 140 signals are expected, approximately 100 are observed, and subsequent analysis revealed that several loops were undergoing μs–ms conformational exchange. It is notable that these residues were well ordered in the X-ray crystal structure of the same protein [13], demonstrating that dynamic solution processes with activation barriers comparable to the amount of thermal energy in the sample can often be missed in crystal structures because the crystallization process pushes the protein into a single energy minimum.

How stable is my protein?

A series of 1D ^{1}H or ^{15}N-HSQC spectra recorded on a sample over a period of time can answer this question. Figure 4A shows changes in the ^{15}N-HSQC spectrum

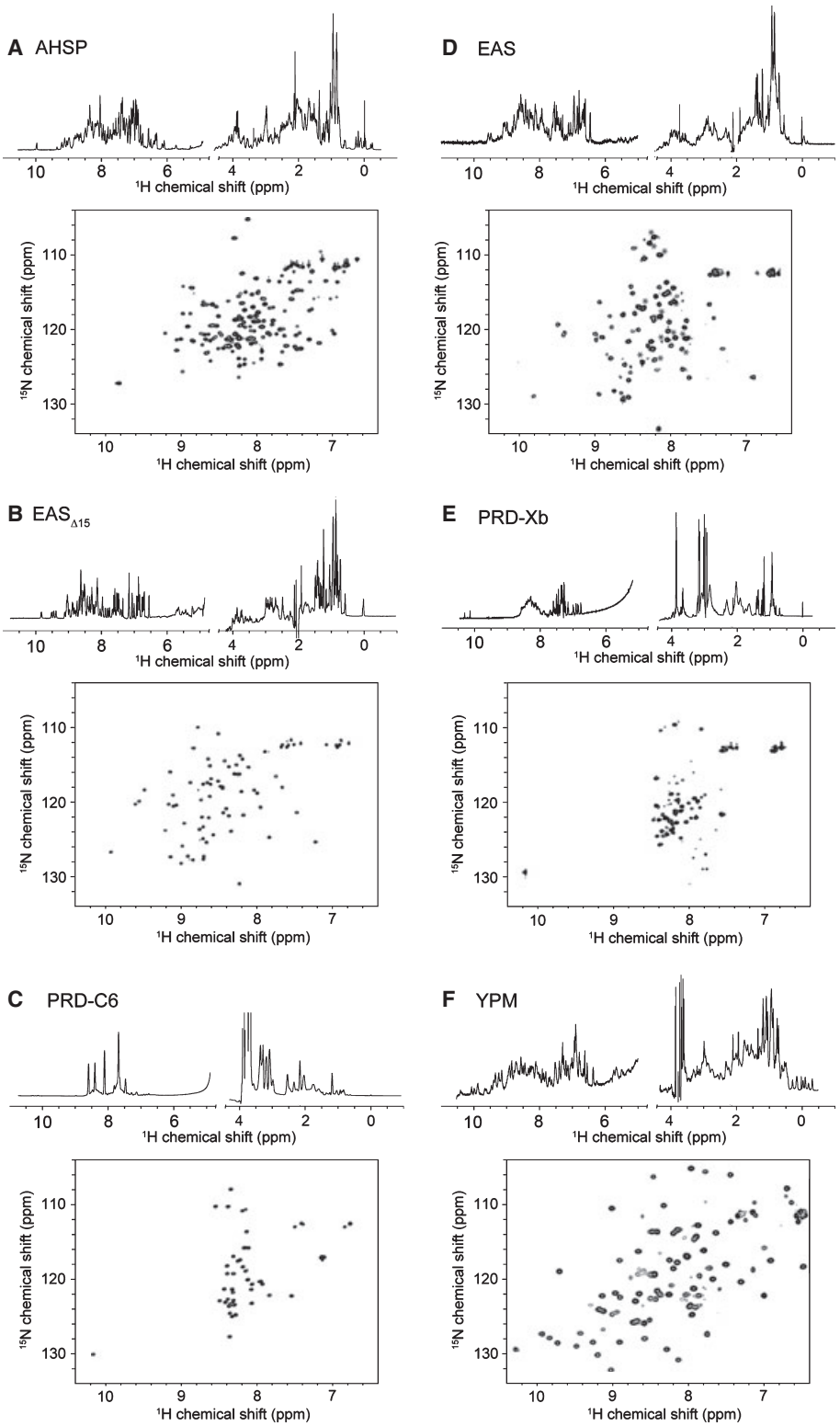

Fig. 3. 1D ^{1}H- and ^{15}N-HSQC spectra of (A) AHSP, a 10 kDa all-α-helical protein; (B) EAS$_{\Delta 15}$, a 7 kDa predominantly β-sheet protein; (C) PRD-C6, a disordered 6 kDa polypeptide; (D) EAS, an 8 kDa predominantly β-sheet protein that contains a 19 residue disordered region; (E) PRD-Xb, a 12 kDa protein segment that exists in a molten globule state; and (F) YPM, a 14 kDa protein for which approximately 25% of the residues are involved in μs–ms dynamics.

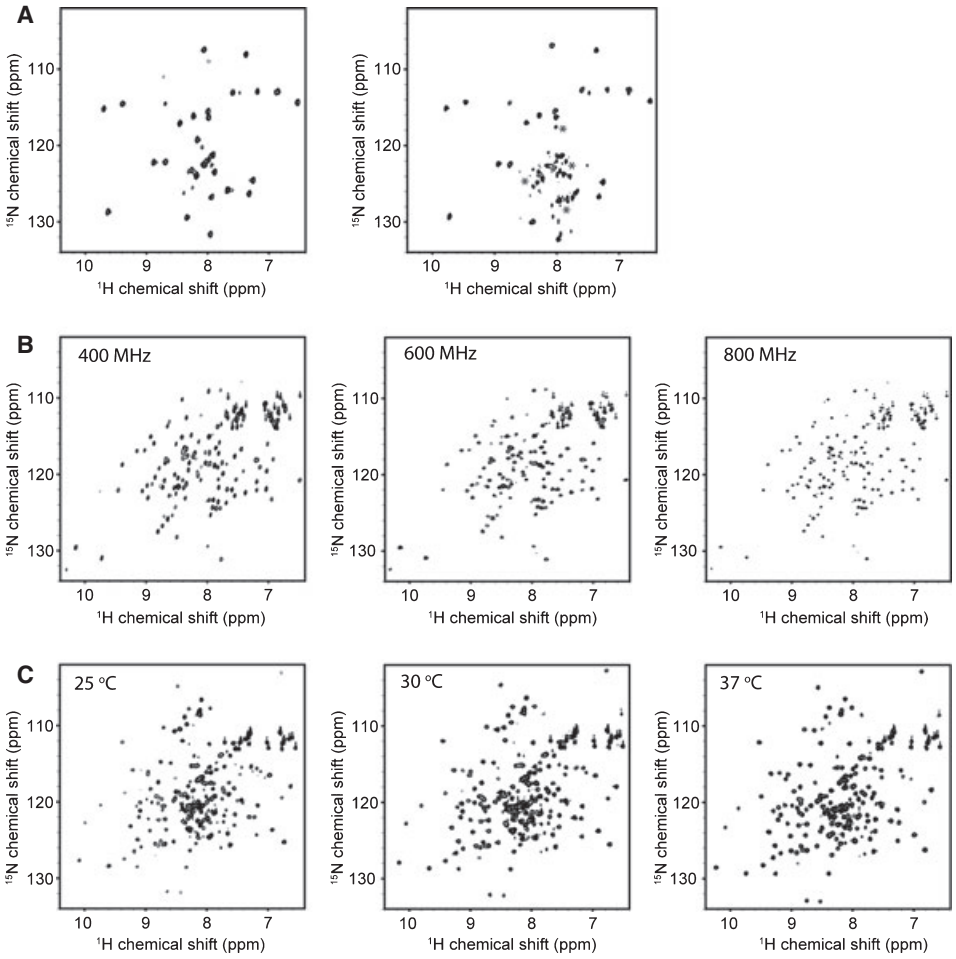

Fig. 4. The effects of various parameters on the appearance of ^{15}N-HSQC spectra. (A) A fresh sample of the MyT1-DNA complex (left) and after 7 days at 25 °C (right). Degradation products are indicated by an asterisk. (B) ^{15}N-HSQC spectra of a 15 kDa protein–peptide complex recorded at 400, 600 and 800 MHz, indicating the improvement in resolution gained from the higher field strength. (C) ^{15}N-HSQC spectra of Flix3 (22 kDa) [62], recorded at 25, 30 and 37 °C, indicating the improvement in spectral quality with increasing temperature. The latter two instruments were equipped with cryoprobes.

of a protein–DNA complex over 1 week. The appearance of a number of new signals in the central part of the spectrum (asterisks) is consistent with either degradation or unfolding of the protein, and suggests that a more stringent purification strategy might be required (i.e. the presence of even very small concentrations of proteases can cause these effects over the long data acquisition periods required for NMR structure determination).

What other parameters affect the appearance of NMR spectra?

The strength of the applied magnetic field has a significant impact on the quality of the recorded spectra. Both sensitivity and resolution are generally improved at higher magnetic field strengths (Fig. 4B). Molecular weight also has a significant influence on NMR linewidths because of the relationship between molecular tumbling and size and, consequently, it is challenging to acquire spectra of proteins bigger than approximately 50 kDa (although see the section 'New Developments' below). For the same reason, macromolecules with extended shapes will also exhibit broader lines than more globular molecules of the same mass.

Changes in temperature can cause a number of effects in spectral appearance. Because higher temperatures cause more rapid tumbling, linewidths can become noticeably narrower, even with a temperature increase of 10 °C. The downside is that many proteins have limited stability at elevated temperatures, and the

rate of exchange of labile amide protons with water is increased, reducing their signal intensity. Temperature changes also alter the rate of other conformational exchange processes, so that, overall, it is always worth screening a range of temperatures before embarking on a detailed NMR study of a protein. Figure 4C shows the ^{15}N-HSQC spectra of a protein for which an increase in temperature gives rise to a substantial improvement in overall spectral quality.

The composition and concentration of buffer components can also affect the quality of the NMR spectrum but, unfortunately, there are no firm guidelines as to which buffers are best for a given protein. A number of additives have been suggested for improving sample stability, including glutamate/arginine mixtures [14], salts such as sodium sulfate, nondenaturing detergents such as triton, and glycerol [15], although it is likely that these will be useful only for a limited subset of proteins. It has long been lamented that there is no simple and rapid buffer screening protocol analogous to the sparse matrix screens employed by X-ray crystallographers. Accordingly, the only way to tell which of a number of sets of buffer conditions will give rise to the best quality NMR spectra is to record those spectra, and this is a lower-throughput process compared to crystallization screening. Automatic NMR sample changers are available, although these are not currently widely used in protein NMR laboratories. The development of an efficient screening process would be a major step forward.

In the analysis of membrane proteins using solution NMR methods, the most significant variable appears to be the choice of solubilizing detergents [16], and a striking example of what can be achieved, namely a ^{15}N-HSQC of the seven-transmembrane-helix G-protein coupled receptor pSRII, is shown in Fig. 5. Nietlispach *et al.* [17] screened a number of detergents, and the spectra obtained from pSRII in diheptanoylphospatidylcholine give spectra that rival those of 'normal' soluble proteins in quality, despite the fact that the protein–micelle complex is approximately 70 kDa in size. This field is likely to expand rapidly over the next few years as our appreciation of the qualities of different detergents improves.

The ease with which 1D ^{1}H and ^{15}N-HSQC spectra can be recorded strongly suggests that these spectra can be routinely recorded by any protein chemist who purifies a protein for structural or biochemical analysis. In most cases, 30–60 min of spectrometer time on a sample at a relatively modest concentration can provide a great deal of insight that cannot be obtained by other methods and thus can inform subsequent experimental design. Once a commitment to the technique is made, however, and a sample is placed into an NMR tube, a whole host of additional possibilities open up. The remainder of this review (as well as the accompanying review [1]) outline the NMR approaches that can be employed to probe the structure, dynamics and function of a macromolecule of interest.

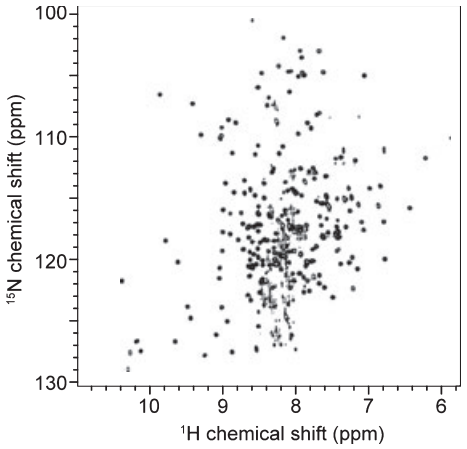

Fig. 5. ^{15}N-HSQC spectrum of the seven-transmembrane-helix G-protein coupled receptor pSRII [17].

Analysis of macromolecular structure by NMR spectroscopy

Introduction

First, what is meant by determining a protein structure? In general, the resolution of an image is defined by the wavelength of the light measured. Thus, to record the image of a molecule, the desired resolution is approximately 0.1 nm (i.e. similar in size to covalent chemical bonds) and the wavelengths required for such measurements are in the X-ray range (0.01–10 nm). Thus, the use of X-ray crystallography allows the measurement of an image of a molecule. In NMR, however, we measure wavelengths in the radiofrequency range (1 mm to 10 km), which is more suitable for imaging elephants than molecules. It is therefore important to remember that an NMR-derived structure is not an image in the sense that an X-ray structure or a picture of your grandmother is. This has advantages and disadvantages. The major advantage is that we can measure much more than just a static image of a molecule; indeed, we often find that a macromolecule does not conform to a single image (e.g. a protein with multiple conformations) or that there is no distinct image at all (e.g. a disordered protein). Moreover, we can study macromolecules in their native solution state rather than in a crystal lattice. On

the downside, much of the life of an NMR structural biologist is spent piecing together indirect evidence of structural features (so-called 'structural restraints') with the aim of reconstructing an image of the macromolecule that is consistent with all of the experimental data (Fig. 6).

How are NMR data used to determine the solution structure of macromolecules? The first task of the NMR spectroscopist is to find the chemical shift of every atom in the molecule, a process referred to as resonance assignment. In the case of proteins, assignments are most commonly made by expressing and purifying uniformly $^{15}N/^{13}C$-labelled protein and recording and analyzing a series of so-called triple resonance NMR experiments [18]. These experiments make connections between the ^{1}H, ^{13}C and ^{15}N nuclei (see below) and the patterns of connections can be mapped onto the protein sequence. Once the chemical shifts of as many atoms as possible have been assigned (typically > 90%), we are ready to start gathering structural restraints. Traditionally, these comprise proton–proton distances, dihedral angles and hydrogen bonds (Fig. 6).

Internuclear interactions and structural restraints

The use of NMR data to determine macromolecular structures relies on the existence (to a first approximation) of two types of interactions between pairs of nuclei that are manifested in NMR spectra. The first of these interactions is the dipolar interaction, particularly between protons. Each proton can sense the presence of other protons that are up to approximately 6 Å away in space and this interaction is measured as a ^{1}H, ^{1}H nuclear Overhauser effect (NOE) in 2D NOESY experiments. For proteins that can be isotopically labelled with ^{13}C and ^{15}N, 3D versions of this experiment are often acquired in which the NOEs are spread (or 'edited') into a third chemical shift dimension (either ^{13}C or ^{15}N), which provides higher spectral resolution and therefore less ambiguity in the NOE assignments.

^{1}H, ^{1}H NOEs are the most important source of structural information in NMR because they provide an indirect measure of the distances between the chemically abundant hydrogen nuclei; pairs of protons that are closer in space give rise to larger NOEs. NOEs are the only NMR-derived structural restraints that, if used

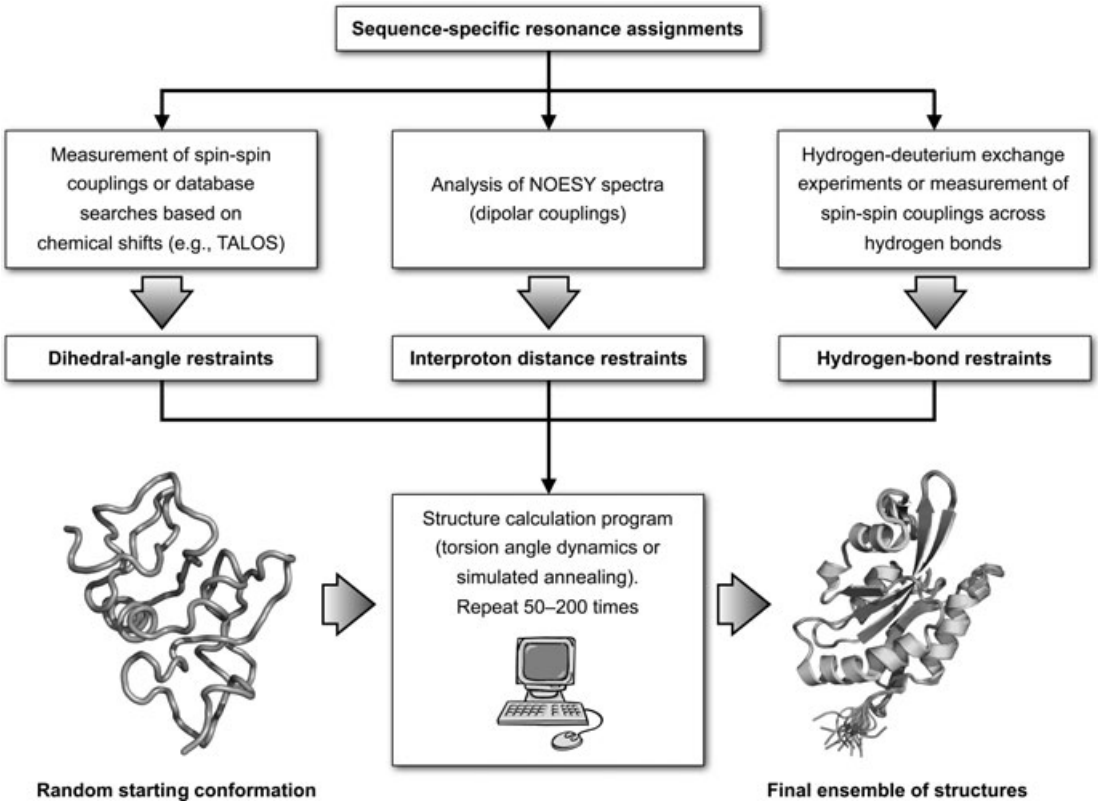

Fig. 6. Overview of the process of macromolecular structure determination using NMR spectroscopy. Analysis of multidimensional NMR spectra leads to three primary sets of structural restraints (interproton distances, dihedral angles and hydrogen bonds) that are used as input to a computer algorithm to reconstruct an image of the molecule.

without any other restraints, would still be capable of routinely producing a reliable high-resolution structure. For even a modest-sized protein of 100 residues, one would expect to measure several thousand distances from NOE data (Fig. 7). Incorrect NOE assignments are usually apparent very early in the structure determination process because they will be inconsistent with the large network of other restraints. Thus, NMR is less prone to the types of major errors that can occur using X-ray crystallography, such as tracing the polypeptide chain backwards in an electron density map [19] or fitting to a mirror image of the map [20].

The second essential interaction is manifested between pairs of nuclei that are close in the covalent structure of the molecule (separated by less than three of four covalent bonds). These scalar (or *J*) couplings are only observed within a residue or between nuclei in adjacent residues, and it is because of this property that so-called triple resonance spectra (which comprise ^{1}H, ^{13}C and ^{15}N frequency dimensions) can be used to unambiguously assign each NMR signal to a particular nucleus in the protein. Information encoded in the excited state of a nucleus (also referred to as coherence or magnetization) can be transferred from one nucleus to the next (e.g. from a ^{15}N nucleus to a $^{13}C_{\alpha}$) via these couplings, establishing connections between the nuclei. The magnitude of these scalar couplings is also a useful parameter; scalar couplings between nuclei that are separated by three covalent bonds vary in a predictable way depending on the dihedral angle about the bond connecting the nuclei [21]. Thus, scalar coupling measurements provide additional structural constraints, particularly for the backbone ϕ angles. In addition, both ϕ and ψ backbone dihedral angles can be robustly estimated based on the correlation between backbone conformation and the chemical shifts of the $^{1}H_{\alpha}$, $^{13}C'$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and backbone ^{15}N nuclei [22,23].

Hydrogen bonds can also be inferred from NMR data and they are useful structural restraints. The rate of exchange of the backbone amide protons with solvent water molecules can be reduced by many orders of magnitude in folded proteins compared to unstructured peptides, largely as a result of hydrogen bond formation. Qualitative analysis of the exchange rate for each amide proton when the solvent is exchanged from $^{1}H_2O$ to $^{2}H_2O$ (also known as D_2O or 'heavy water') allows slowly-exchanging protons to be identified. Note that this approach does not reveal the identity of the hydrogen bond acceptor, which has to be inferred from preliminary structure calculations. More recently, scalar couplings have been measured across hydrogen bonds in both proteins [24–28] and nucleic acids [29,30]. This approach has the advantage of identifying both the donor and the acceptor atoms, although, unfortunately, the couplings are very small in proteins and therefore difficult to measure [31,32].

How are the various structural restraints used to calculate a structure?

The final step in protein structure determination using NMR is to use computer software that combines all of the NMR-derived conformational restraints with additional restraints based on the covalent structure of the protein (i.e. bond lengths and bond angles) and known atomic properties (i.e. atomic radius, mass, partial

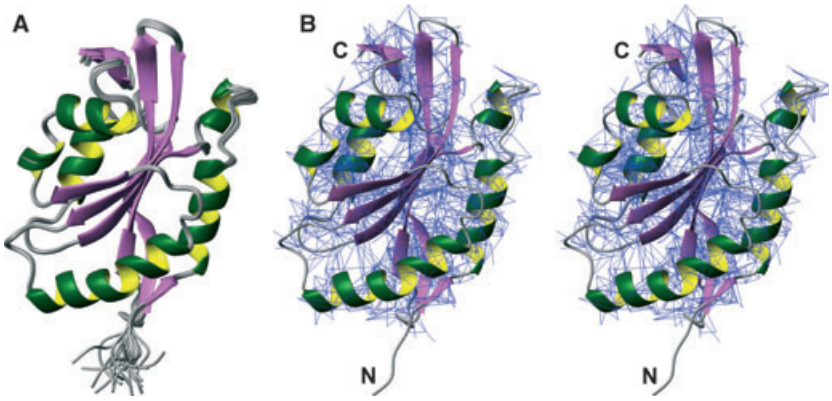

Fig. 7. (A) An overlay of the ensemble of 20 structures of chicken cofilin (PDB coordinate file: 1TVJ) optimized for lowest backbone rmsd over residues 5–166 of the mean coordinate structure; this superposition yielded an rmsd of 0.25 ± 0.05 Å [63]. (B) Stereoview of the first structure from the same ensemble showing the network of interproton distance restraints that was used in the structure calculations; each blue line represents a separate restraint. Note the absence of NOESY-derived distance restraints for the four N-terminal residues; this explains the poor overlay obtained for this part of the structure and suggests that these residues are highly dynamic in solution. Consistent with this hypothesis, Ser3 is a target for phosphorylation by LIM kinase [63].

charge, etc.) to calculate a 3D structure that is consistent with all of the restraints (Fig. 6). The primary experimental restraints are interproton distances derived from NOESY cross-peak intensities, dihedral-angle restraints derived from either coupling constants or database searches based on chemical shift information, as well as hydrogen-bond restraints. It is the quantity rather than the precision of these restraints that is important for NMR structure determination [33,34]. In mathematical parlance, we aim to collect so many restraints that the problem (i.e. determination of a unique 3D structure) is overdetermined. Hence, it is common in NMR structure calculations to conservatively set the restraints and their associated errors because over-restraining the distances and angle estimates is more likely to lead to errors.

Although the first protein structure determined using NMR was reported in 1985 [35], unfortunately, there is still no consensus method for deriving a 3D structure from NMR-derived conformational restraints. In general, however, most of the available software packages use a similar strategy, namely molecular dynamics simulations in the presence of the experimental constraints derived from the NMR data (restrained molecular dynamics or RMD). In classical molecular dynamics simulations, Newton's equations of motion are solved for all atoms under the influence of an empirically-derived physical force field [36]. The RMD strategy adds restraining potentials to the force field so that the structure can be refined against terms describing covalent geometry, nonbonded interactions (i.e. V_{physical}) and the experimentally-derived distance ($V_{\text{distances}}$) and dihedral-angle (V_{dihedral}) restraints. Thus, the overall force field can be represented as:

$$V_{\text{total}} = V_{\text{physical}} + V_{\text{distances}} + V_{\text{dihedral}}$$

In these calculations, the motion of the molecule is simulated for sufficient time to allow sampling of large regions of conformational space with the aim of converging on the structure with the global energy minimum by the end of the simulation. Early stages of the calculations are carried out at high temperature (so that the atoms have high kinetic energy), thereby maximizing the sampling of conformational space and reducing the chance of the protein getting trapped in a 'dead-end' conformation. As the calculation proceeds, the temperature is reduced so that the protein ends up in an energy minimum (hopefully, the global minimum) corresponding to a structure with good covalent geometry, favourable nonbonded interactions and minimal violations of the experimental constraints.

An alternative approach to RMD is torsion angle dynamics in which the molecular dynamics simulation is performed by solving Lagrange's equations of motion with torsion angles as degrees of freedom. Working in torsion angle space reduces the degrees of freedom by approximately ten-fold compared to simulations in Cartesian coordinate space because the parameters defining the covalent geometry are kept fixed at their optimal values. Thus, torsion angle dynamics, as implemented in software such as CYANA [37], is computationally much faster than classical RMD.

Note that there is a fundamental difference between molecular dynamics simulations used to calculate NMR structures and those that aim to simulate the 'real-life' dynamics of a biomolecular system. In the former case, the trajectory of the system is unimportant and probably bears little resemblance to the real solution dynamics of the macromolecule; the aim is simply to compute as efficiently as possible a stereochemically correct structure that satisfies all of the experimentally-derived structural restraints.

How does one look at an NMR structure?

X-ray crystallography leads to a single static image of a macromolecule. Those regions of a protein or DNA/RNA molecule that are flexible in the crystal do not provide coherent X-ray scattering and hence do not contribute to the final electron density map. Thus, for all intents and purposes, they can effectively be ignored. NMR structure determination leads to a very different 'picture' of a macromolecule. First, all regions of a protein or DNA/RNA molecule will be observed in NMR spectra (unless they are undergoing μs–ms exchange), even those regions that are very mobile. All segments of the protein will therefore appear in the final pictorial representation of the structure, even if their conformation and stereochemistry are poorly defined. The fact that disordered regions are visible in NMR spectra can be helpful when assessing the binding of a protein to potential ligands because disordered segments can often mediate protein interactions. Second, NMR structure determination does not lead to a single 'image'. Rather, the structure calculation process is repeated many times, each time starting from a randomly generated structure, aiming to generate an ensemble of structures, each of which satisfies the input experimental restraints.

The ensemble is usually displayed as an overlay of the individual members to provide the lowest rmsd of individual structures from the mean structure. Because the backbone of a protein is more rigid than the side chains, the rmsd is usually calculated over only the backbone heavy atoms (N, C_α, C′). For structures in which there are no disordered or highly dynamic

regions, most or all residues will be included in the rmsd calculation (Fig. 7A). However, regions of a protein that are highly flexible will access multiple conformations during the time course of an NMR experiment and, often, none of these conformations will be maintained for sufficient time to yield representative NOEs. Thus, there will be few or no NOE-derived interproton distance restraints for these regions and their conformation will differ in each member of the ensemble. These regions are excluded from the rmsd calculation. As an example, Fig. 8 shows the NMR structure of a protein with a highly flexible C-terminal region. An overlay of the ensemble of structures over all 45 residues yields a rather uninformative 'furball' (Fig. 8A). In comparison, an overlay over only the structured region of the protein (residues 3–32) reveals a compact, well-ordered core and disordered N- and C-termini (Fig. 8B).

How does one interpret poorly overlaid regions such as the N-and C-termini in Figure 8B? So long as the NMR structure determination has been performed competently and there are no errors in the sequence-specific resonance assignments (e.g. as a result of exchange or severe spectral overlap), a poor overlay indicates that these regions of the protein are highly mobile in solution. One of the advantages of NMR is that additional so-called 'relaxation' experiments can then be performed to probe the dynamics of these regions, as discussed in the accompanying review [1].

How good is an NMR-derived structure?

The rmsd value for an ensemble provides a measure of the precision (but not the accuracy) of the structures. In general, a well defined NMR structure should have a backbone rmsd < 0.5 Å and an all-heavy-atom rmsd < 1.0 Å, measured over the structured part of the protein. Those wishing to use an NMR structure for structure-based drug design or ligand docking should use extreme caution if the rmsd is higher than these values.

Measurement of the accuracy of NMR-derived structures is a much more difficult task than estimating their precision. An absolute measure of the accuracy of an NMR-derived structure is not possible in the absence of any knowledge about the 'true' structure and therefore it has to be measured by some statistic [38]. The most reliable indicator of the quality of an NMR-derived structure is its stereochemical merit, as judged by software such as PROCHECK-NMR [39], WHATIF [40] and MOLPROBITY [41], which report numerous measures of stereochemical merit, including Ramachandran plot quality, deviations of bond lengths, bond angles and dihedral angles from ideality, unfavourable sidechain rotamers, and bad nonbonded interactions. MOLPROBITY additionally offers all-atom contact analysis and provides an overall score that allows the structure to be ranked on a percentile basis against other structures in the PDB. A MOLPROBITY score that caused a structure to be ranked in the bottom 20th percentile or lower would be cause for concern, and should provoke a detailed analysis of the MOLPROBITY output.

A word of caution, however, is warranted when using these software packages. By contrast to X-ray crystallography, where highly dynamic regions of the protein do not appear in the electron density maps and thus are omitted from the final coordinate file, all regions of the protein are modelled in NMR structure calculations. As discussed above, highly dynamic regions of the protein, in which multiple conformations are accessed during the timescale of the NMR experiment, will either have a completely ill defined

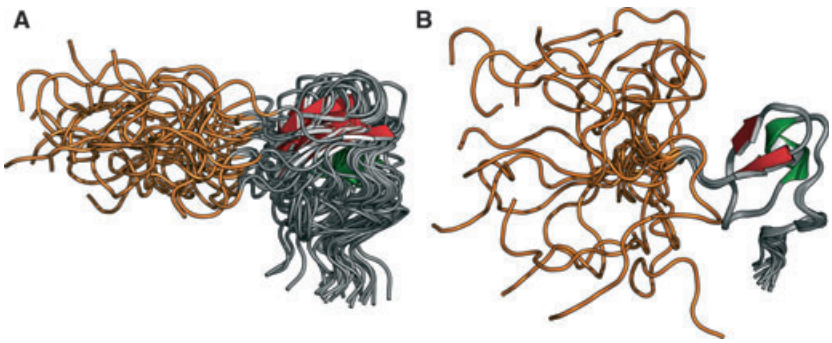

Fig. 8. Ensemble of 20 NMR-derived structures of ω-ACTX-Hv2a, a specific blocker of insect voltage-gated calcium channels [64]. (A) In this view, the structures have been overlaid to minimize the backbone rmsd over all 45 residues; (B) An overlay over the backbone atoms of residues 3–32 only reveals that the furball in part (A) results from inclusion of the highly disordered C-terminal region in the rmsd calculation. When this region is excluded, the structured N-terminal core, which includes three disulfide bonds (not shown), is clearly visible.

conformation as a result of the lack of NOE information or else an unrealistic one as a result of an averaging of the NOEs and coupling constants. Thus, these regions of the protein are likely to have poor Ramachandran plot quality and bad side-chain rotamer distributions, although these analyses are meaningless when applied to such mobile regions. Inclusion of such regions in a MOLPROBITY, PROCHECK or WHATIF analysis may therefore give a false indication of the quality of the well structured region of the protein or peptide. Thus, these regions should be omitted from the stereochemical analysis, just as they effectively are in analysis of X-ray crystal structures.

Because NMR structures are not images, it is not possible to quantitatively define the resolution. A recent study of packing quality for all PDB structures suggested that packing correlated closely with resolution for X-ray structures, and concluded that the distribution of packing quality for NMR structures resembled that of 3 Å X-ray structures [42]. However, a number of problems exist with this analysis, not least of which is the fact that packing forces will be greater in a crystal lattice and thus may well not reflect the situation in solution. Instead, from both an overall comparison with X-ray structures and PROCHECK analysis of 'equivalent resolution', we can infer that most high-quality NMR structures have an equivalent resolution of 2–3 Å. NMR structures are often qualitatively defined as high-, medium- or low-resolution based on a panel of quantitative measures of precision and stereochemical quality such as rmsd, Ramachandran plot quality and the number of experimentally-derived structural restraints per structured residue. Table 1 provides a guide for judging the quality of NMR structures.

Because NMR structure determination typically leads to an ensemble of 20–25 structures, a commonly asked question is what structure or structures from the ensemble should be used for applications such as drug design, docking and homology modelling? Although 'average structures' were commonly calculated from the ensemble in the past, these are by no means guaranteed to be of higher quality than the individual conformers. If the downstream application has to be restricted to a single structure from the ensemble, then, in almost all cases, one should use the first structure from the NMR ensemble. When NMR ensembles are deposited in the PDB, the submitters typically add each of the structures into the coordinate file in order of their perceived quality (i.e. from highest to lowest quality), based on the output of the structure determination or post-calculation analysis software. However, if computational power is not limiting, then we recommend performing applications such as docking, structure-based drug design and *in silico* screening using each member of the ensemble to take into account any inherent flexibility in binding sites.

New developments

Macromolecular NMR is getting bigger

Structure determination using NMR spectroscopy is generally restricted to macromolecules smaller than approximately 25 kDa. This is readily apparent by examining the percentage of structures determined

Table 1. A guide for judging the 'resolution' of NMR-derived protein structures.

Assessment criterion	Very high resolution	High resolution	Medium resolution	Low resolution
Restraints per residue[a]	> 18	14–18	10–15	< 10
Backbone rmsd (Å)[b]	< 0.3	0.3–0.5	0.5–0.8	> 0.8
Heavy-atom rmsd (Å)[b]	< 0.75	0.75–1.0	1.0–1.5	> 1.5
Ramachandran				
Plot quality (%)[c]	> 95	85–95	75–85	< 75
Example PDB file	1TVJ [63]	2IL8 [65]	2FE0 [66]	1LMM [67]

[a] Total number of interproton-distance, dihedral-angle and hydrogen-bond restraints per residue. Disordered regions should be excluded from this calculation, and it is important that only structurally relevant restraints are included in the count. Unfortunately, many NMR studies give a misleading indication of the true number of structural restraints by including interproton distances that do not restrain the protein conformation. For example, an upper-limit distance restraint of 4.5 Å between the H_α of residue *i* and H_N of residue *i + 1* is not a structural restraint because this distance is always less than 3.5 Å, regardless of the conformation of the protein [68]. Note that interproton distance restraints are often divided into categories of 'intraresidue', 'sequential' (NOEs between protons on adjacent residues), 'medium range' (NOEs between protons separated by two to five residues) and 'long range' (NOEs between protons separated by more than residues). The number of medium-range and long-range restraints is the most important factor when determining the global fold of the protein. [b] rmsd calculated versus mean coordinate structure, with disordered regions excluded. [c] Percentage of residues in most favoured region of the Ramachandran plot as judged by MOLPROBITY. Note that these numbers will be slightly lower if PROCHECK is used for stereochemical analysis because of the slightly different way in which the most favoured regions of the Ramachandran plot are defined.

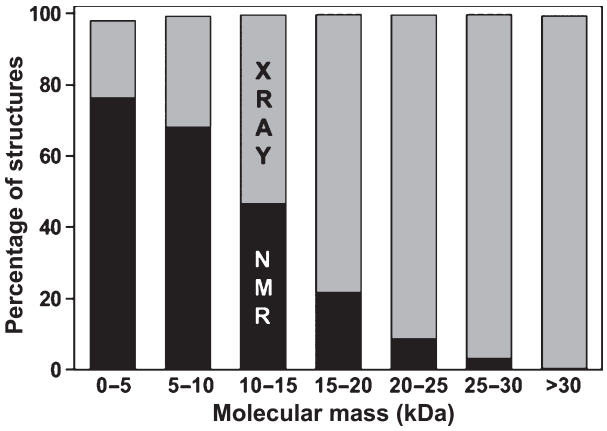

Fig. 9. Histogram comparing the percentage of protein structures in the PDB determined using solution-state NMR spectroscopy (black bars) and X-ray crystallography (grey bars). NMR dominates the PDB for small proteins, whereas X-ray crystallography is dominant for proteins > 15 kDa.

using NMR as a function of molecular mass (Fig. 9). Although NMR dominates the PDB for proteins smaller than 10 kDa, the vast majority of structures determined for proteins > 30 kDa have been solved using X-ray crystallography. The upper limit of 25 kDa for routine NMR structure determination might be considered quite restrictive, although it includes numerous small proteins as well as most protein domains, which have an average size of approximately 17 kDa [43].

The 25 kDa cap arises because the NMR excited state becomes more short-lived as the molecular size increases as a result of larger molecules tumbling more slowly in solution (technically speaking, their molecular correlation time increases). Consequently, the transfer of magnetization through scalar couplings becomes less efficient, thereby resulting in poor quality triple resonance spectra that prevent chemical shift assignments from being made. It was demonstrated in the late 1990s that by 'throwing away' some components of the magnetization, and retaining the fraction that decays most slowly, improved spectra for large proteins can be acquired. However, there are two caveats to this approach, which is known as transverse relaxation optimized spectroscopy (TROSY) [44]. First, it is only suitable for NMR spectrometers operating at frequencies > 700 MHz and, second, the protein must be labelled with ^{2}H atoms (which further reduces signal decay). TROSY-based methods were used to solve the solution structure of malate synthase G, an 82 kDa enzyme, setting an NMR size record that will be difficult to routinely match in the near future [45].

TROSY-based resonance assignment is very demanding and, in most cases, it is advisable to exhaust the X-ray crystallography approach before attempting NMR structural studies of proteins larger than 30 kDa. However, TROSY-based experiments can provide a convenient route for monitoring binding interfaces on small proteins (< 25 kDa) as they form larger complexes through interactions with, for example, another protein, a lipid membrane or RNA/DNA. This approach has been used to study protein–protein interactions in complexes as large as 870 kDa [46,47], and is discussed in more detail in the accompanying review [1].

Increasing the diversity of experimental restraints

Two additional classes of structural restraints have recently been added to the toolbox used by NMR spectroscopists to study macromolecules. Although not widely used at present, these techniques are likely to become more prevalent as NMR spectroscopists tackle larger and more complex macromolecular systems [48].

Residual dipolar couplings rely on the fact that the dipolar interaction between two nuclei depends on the orientation of the molecule in solution with respect to the spectrometer magnet. This effect is normally averaged to zero by rapid molecular tumbling, although partial alignment of a macromolecule, as achieved by steric restriction using polyacrylamide gels, bacteriophage or other reagents [49,50], can recover the couplings. The magnitude of the coupling for each ^{1}H-^{15}N pair, for example, provides information on the orientation of all N-H bond vectors relative to a single molecular axis. Because these restraints do not provide information on the proximity of the N-H bonds, they cannot be used on their own to determine a high-resolution structure. However, if the protein fold is known, the residual dipolar couplings (RDCs) can be used to either refine a structure [51] or to help orient two distal domains [52]. RDCs can furthermore be used to supplement NOE data. Indeed, RDCs have found application in several areas where long-range NOEs are not abundant either as a result of the scarcity of protons or a lack of tertiary structure; examples include small organic molecules [53], complex carbohydrates [54], DNA [55] and RNA [56].

The introduction of a paramagnetic moiety into a protein (e.g. a nitroxide radical that attaches to cysteines or a lanthanide metal, which can be attached via covalent tags) has two significant effects on an NMR spectrum. First, it greatly broadens the NMR signal of nuclei close to the paramagnetic centre, an effect known as paramagnetic relaxation enhancement. Thus, the acquisition of HSQC spectra in the absence and presence of the paramagnetic species will quickly identify nuclei in the vicinity (i.e. < 30 Å) of the

paramagnetic centre [57]. Second, certain classes of paramagnetic species can affect the chemical shift of nearby nuclei (the pseudo contact shift) [58]. By contrast to RDCs, the pseudo contact shifts provide both distance and angular information, and over quite large distances (≤ 40 Å) [58]. The unique nature of such restraints (in particular in situations where conformational restraints are hard to come by) makes this a very interesting approach that promises to further expand the utility of NMR for probing macromolecular structure [48,59].

Faster is better

The NMR structure determination process can be divided into four distinct steps: (i) data acquisition and processing; (ii) data extraction; (iii) resonance assignment; and (iv) structure calculation. Methods for increasing the throughput of each of these steps are currently being developed. Fast data acquisition methods can reduce the time required to acquire a set of triple resonance data from 2 to 4 weeks to only a few days. These methods do not change the types of experimentally-derived structural restraints that are used to calculate structures, nor the method of structure calculation; they simply speed up the rate of data acquisition. Using these approaches in combination with automated spectral assignment and structure calculation (see below), it is now possible to use NMR to determine a high-quality protein structure in less than 1 week [60].

A number of algorithms have also been designed to analyze triple resonance and NOESY spectra with minimal or no user input. For example, the PINE server (http://pine.nmrfam.wisc.edu) allows online submission of chemical shift lists from triple resonance spectra together with an e-mail address to which chemical shift assignments are sent when calculations are finished. These methods can be very powerful and, even if not used for complete automated assignment, they can be a useful first pass to facilitate manual assignment. Finally, once resonances have been assigned, the relevant structural restraints must be extracted. Software such as CYANA [37] and ARIA [61] allow the ability to automatically assign NOESY spectra and calculate structures; this dramatically improves the speed of the NMR structure determination process because, particularly for homonuclear NMR data, much more time is usually spent analyzing the data than collecting it. By contrast with the manual approach, which can take weeks or even months, the automated process performed by CYANA takes approximately 1 h on a laptop computer for a protein of approximately 10 kDa, or just few minutes on even a modest server.

In conclusion, we hope that this review has alerted non-NMR spectroscopists to the many potential uses of NMR in the area of structural biology, and will stimulate them to discuss with their local NMR spectroscopist how this technique can help with their own research. Of course, one of the great advantages of NMR spectroscopy is the multitude of biological problems to which it can be applied and, in the accompanying review [1], we discuss some of the major biological applications of NMR beyond macromolecular structure determination.

Acknowledgements

The authors acknowledge financial support from the Queensland Smart State Research Facilities Fund, and Discovery Grants DP0774245, DP1095728 and DP0879121 from the Australian Research Council. We are grateful to Dr Daniel Nietlispach for providing a ^{15}N-HSQC spectrum of pSRII.

References

1 Bieri M, Kwan AH, Mobli M, King GF, MacKay JP & Gooley PR (2011) Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination. *FEBS J* **278**, 704–715.
2 Keeler J (2005) *Understanding NMR Spectroscopy*. Wiley.
3 Levitt MH (2008) *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2 edn. Wiley, Chichester.
4 Cavanagh J, Fairbrother WJ, Palmer III AG, Skelton NJ & Rance M (2007) *Protein NMR Spectroscopy*, 2nd Edn. Academic Press, San Diego.
5 Bodenhausen G & Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett* **69**, 185–188.
6 Flynn PF, Mattiello DL, Hill HDW & Wand AJ (2000) Optimal use of cryogenic probe technology in NMR studies of proteins. *J Am Chem Soc* **122**, 4823–4824.
7 Adler AJ, Greenfield NJ & Fasman GD (1973) Circular dichroism and optical rotatory dispersion of proteins and polypeptides. *Methods Enzymol* **27**, 675–735.
8 Takahashi H & Shimada I (2009) Production of isotopically labeled heterologous proteins in non-E. coli prokaryotic and eukaryotic cells. *J Biomol NMR* **46**, 3–10.
9 Coen M, Holmes E, Lindon JC & Nicholson JK (2008) NMR-based metabolic profiling and metabonomic approaches to problems in molecular toxicology. *Chem Res Toxicol* **21**, 9–27.
10 King GF & Kuchel PW (1994) Theoretical and practical aspects of NMR studies of cells. *Immunomethods* **4**, 85–97.

11 Lesage A (2009) Recent advances in solid-state NMR spectroscopy of spin I = 1/2 nuclei. *Phys Chem Chem Phys* **11**, 6876–6891.
12 Ptitsyn OB (1995) Molten globule and protein folding. *Adv Protein Chem* **47**, 83–229.
13 Donadini R, Liew CW, Kwan AH, Mackay JP & Fields BA (2004) Crystal and solution structures of a superantigen from *Yersinia pseudotuberculosis* reveal a jelly-roll fold. *Structure* **12**, 145–156.
14 Hautberguea GM & Golovanov AP (2007) Increasing the sensitivity of cryoprobe protein NMR experiments by using the sole low-conductivity arginine glutamate salt. *J Magn Res* **191**, 335–339.
15 Ducat T, Declerck N, Gostan T, Kochoyan M & Demene H (2006) Rapid determination of protein solubility and stability conditions for NMR studies using incomplete factorial design. *J Biomol NMR* **34**, 137–151.
16 Page RC, Moore JD, Nguyen HB, Sharma M, Chase R, Gao FP, Mobley CK, Sanders CR, Ma L, Sonnichsen FD *et al.* (2006) Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins. *J Struct Func Genom* **7**, 51–64.
17 Gautier A, Kirkpatrick JP & Nietlispach D (2008) Solution-state NMR spectroscopy of a seven-helix transmembrane protein receptor: backbone assignment, secondary structure, and dynamics. *Angew Chem Int Ed* **47**, 7297–7300.
18 Sattler M, Schleucher J & Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* **34**, 93–158.
19 Janin J (1990) Errors in three dimensions. *Biochimie* **72**, 705–709.
20 Chang G, Roth CB, Reyes CL, Pornillos O, Chen YJ & Chen AP (2006) Retraction. *Science* **314**, 1875.
21 Karplus M (1963) Vicinal proton coupling in nuclear magnetic resonance. *J Am Chem Soc* **85**, 2870–2871.
22 Cornilescu G, Delaglio F & Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289–302.
23 Shen Y, Delaglio F, Cornilescu G & Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213–223.
24 Blake PR, Park JB, Adams MWW & Summers MF (1992) Novel observation of NH•••S(Cys) hydrogen-bond-mediated scalar coupling in ^{113}Cd-substituted rubredoxin from *Pyrococcus furiosus*. *J Am Chem Soc* **114**, 4931–4933.
25 Blake PR, Lee B, Summers MF, Adams MW, Park JB, Zhou ZH & Bax A (1992) Quantitative measurement of small through-hydrogen-bond and 'through-space' ^{1}H-^{113}Cd and ^{1}H-^{199}Hg *J* couplings in metal-substituted rubredoxin from *Pyrococcus furiosus*. *J Biomol NMR* **2**, 527–533.
26 Cordier F & Grzesiek S (1999) Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. *J Am Chem Soc* **121**, 1601–1602.
27 Cordier F, Rogowski M, Grzesiek S & Bax A (1999) Observation of through-hydrogen-bond $^{2h}J_{HC'}$ in a perdeuterated protein. *J Magn Reson* **140**, 510–512.
28 Cornilescu G, Hu J-S & Bax A (1999) Identification of the hydrogen bonding network in a protein by scalar couplings. *J Am Chem Soc* **121**, 2949–2950.
29 Dingley AJ & Grzesiek S (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^{2}J_{NN}$ couplings. *J Am Chem Soc* **120**, 8293–8297.
30 Pervushin K, Ono A, Fernández C, Szyperski T, Kainosho M & Wüthrich K (1998) NMR scalar couplings across Watson-Crick base pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. *Proc Natl Acad Sci USA* **95**, 14147–14151.
31 Grzesiek S, Cordier F & Dingley AJ (2001) Scalar couplings across hydrogen bonds. *Methods Enzymol* **338**, 111–133.
32 Dingley AJ, Cordier F, Jaravine VA & Grzesiek S (2003) Scalar couplings across hydrogen bonds. In *BioNMR in Drug Research* (Zerbe O ed), pp 207–226. Wiley-VCH, Weinheim.
33 Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York.
34 Clore GM, Robien MA & Gronenborn AM (1993) Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. *J Mol Biol* **231**, 82–102.
35 Williamson MP, Havel TF & Wüthrich K (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J Mol Biol* **182**, 295–315.
36 van Gunsteren WF & Berendsen HJC (1990) Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew Chem Int Ed* **29**, 992–1023.
37 Güntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353–378.
38 Brunger AT, Clore GM, Gronenborn AM, Saffrich R & Nilges M (1993) Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. *Science* **261**, 328–331.
39 Laskowski RA, MacArthur MW, Moss DS & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structure coordinates. *J Appl Crystallogr* **26**, 283–291.
40 Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8**, 52–56.
41 Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB III,

Snoeyink J, Richardson JS *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, 375–383.
42 Sheffler W & Baker D (2009) RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci* **18**, 229–239.
43 Shen M-Y, Davis FP & Sali A (2005) The optimal size of a globular protein domain: a simple sphere-packing model. *Chem Phys Lett* **405**, 224–228.
44 Pervushin K, Riek R, Wider G & Wüthrich K (1997) Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* **94**, 12366–12371.
45 Tugarinov V, Choy WY, Orekhov VY & Kay LE (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc Natl Acad Sci USA* **102**, 622–627.
46 Fiaux J, Bertelsen EB, Horwich AL & Wuthrich K (2002) NMR analysis of a 900K GroEL GroES complex. *Nature* **418**, 207–211.
47 Pellecchia M, Sebbel P, Hermanns U, Wüthrich K & Glockshuber R (1999) Pilus chaperone FimC-adhesin FimH interactions mapped by TROSY-NMR. *Nat Struct Mol Biol* **6**, 336–339.
48 Bertini I, Luchinat C & Parigi G (2002) Paramagnetic constraints: an aid for quick solution structure determination of paramagnetic metalloproteins. *Concepts Magn Reson* **14**, 259–286.
49 Prestegard JH, Bougault CM & Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. *Chem Rev* **104**, 3519–3540.
50 Kummerlöwe G & Luy B (2009) Residual dipolar couplings as a tool in determining the structure of organic molecules. *Trends Analyt Chem* **28**, 483–493.
51 Tjandra N & Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114.
52 Fushman D, Varadan R, Assfalg M & Walker O (2004) Determining domain orientation in macromolecules by using spin-relaxation and residual dipolar coupling measurements. *Prog Nucl Magn Reson Spectrosc* **44**, 189–214.
53 Thiele CM (2008) Residual dipolar couplings (RDCs) in organic structure determination. *Eur J Org Chem* **2008**, 5673–5685.
54 Xia J & Margulis C (2008) A tool for the prediction of structures of complex sugars. *J Biomol NMR* **42**, 241–256.
55 Vermeulen A, Zhou H & Pardi A (2000) Determining DNA global structure and DNA bending by application of NMR residual dipolar couplings. *J Am Chem Soc* **122**, 9638–9647.
56 Al-Hashimi HM, Gorin A, Majumdar A, Gosser Y & Patel DJ (2002) Towards structural genomics of RNA: rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings. *J Mol Biol* **318**, 637–649.
57 Battiste JL & Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry* **39**, 5355–5365.
58 Allegrozzi M, Bertini I, Janik MBL, Lee Y-M, Liu G & Luchinat C (2000) Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion. *J Am Chem Soc* **122**, 4154–4161.
59 Otting G (2008) Prospects for lathanides in structural biology by NMR. *J Biomol NMR* **42**, 1–9.
60 Liu G, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A *et al.* (2005) NMR data collection and analysis protocol for high-throughput protein structure determination. *Proc Natl Acad Sci USA* **102**, 10487–10492.
61 Habeck M, Rieping W, Linge JP & Nilges M (2004) NOE assignment with ARIA 2.0: the nuts and bolts. *Methods Mol Biol* **278**, 379–402.
62 Bhati M, Lee C, Nancarrow AL, Lee M, Craig VJ, Bach I, Guss JM, Mackay JP & Matthews JM (2008) Implementing the LIM code: the structural basis for cell type-specific assembly of LIM-homeodomain complexes. *EMBO J* **27**, 2018–2029.
63 Gorbatyuk VY, Nosworthy NJ, Robson SA, Bains NPS, Maciejewski MW, dos Remedios CG & King GF (2006) Mapping the phosphoinositide-binding site on chick cofilin explains how PIP_2 regulates the cofilin-actin interaction. *Mol Cell* **24**, 511–522.
64 Wang XH, Connor M, Wilson D, Wilson HI, Nicholson GM, Smith R, Shaw D, Mackay JP, Alewood PF, Christie MJ *et al.* (2001) Discovery and structure of a potent and highly specific blocker of insect calcium channels. *J Biol Chem* **276**, 40306–40312.
65 Clore GM, Appella E, Yamada M, Matsushima K & Gronenborn AM (1990) Three-dimensional structure of interleukin 8 in solution. *Biochemistry* **29**, 1689–1696.
66 Tull D, Naderer T, Spurck T, Mertens HD, Heng J, McFadden GI, Gooley PR & McConville MJ (2010) Membrane protein SMP-1 is required for normal flagellum function in Leishmania. *J Cell Sci* **123**, 544–554.
67 Escoubas P, Bernard C, Lambeau G, Lazdunski M & Darbon H (2003) Recombinant production and solution structure of PcTx1, the specific peptide inhibitor of ASIC1a proton-gated cation channels. *Protein Sci* **12**, 1332–1343.
68 Billeter M, Braun W & Wüthrich K (1982) Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Computation of sterically allowed proton–proton distances and statistical analysis of proton–proton distances in single crystal protein conformations. *J Mol Biol* **155**, 321–346.