NMR structure determination of proteins and protein complexes larger than 20 kDa

G Marius Clore and Angela M Gronenborn

Recent advances in multidimensional nuclear magnetic resonance methodology to obtain ^{1}H, ^{15}N and ^{13}C resonance assignments, interproton distance and torsion angle restraints, and restraints that characterize long-range order, coupled with new methods of structure refinement and novel methods for reducing linewidths, have permitted three-dimensional solution structures of single chain proteins in excess of 250 residues and multimeric proteins in excess of 40 kDa to be solved. These developments may permit the determination by nuclear magnetic resonance of macromolecular structures up to molecular weights in the 50–60 kDa range, thereby bringing into reach numerous systems of considerable biological interest, including a large variety of protein–protein and protein–nucleic acid complexes.

Address
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, USA

Current Opinion in Chemical Biology 1998, **2**:564–570

http://biomednet.com/elecref/1367593100200564

Abbreviations

2D	two-dimensional
3D	three-dimensional
4D	four-dimensional
NOE	nuclear Overhauser enhancement
SIV	simian immunodeficiency virus
T_1	longitudinal relaxation time
T_2	transverse relaxation time
TROSY	transverse relaxation-optimized spectroscopy

Introduction

The size of macromolecular structures that can be solved by nuclear magnetic resonance (NMR) has been dramatically increased over the past few years [1]. The development of a wide range of two-dimensional (2D) NMR experiments (using ^{1}H active nuclei) in the early 1980s culminated in the determination of the structures of a number of small proteins [2,3]. Under exceptional circumstances, 2D NMR techniques can be applied successfully to the structure determination of proteins up to ~100 residues [4,5]. Beyond 100 residues (~10 kDa), however, 2D NMR methods tend to fail, principally due to spectral complexity that cannot be resolved in two dimensions. In the late 1980s and early 1990s, a series of major advances took place with the introduction of experiments that significantly extended the spectral resolution by extending the dimensionality from two to three and four dimensions by using ^{13}C and ^{15}N active nuclei (see [1] for a review). In addition, the combination of such multidimensional experiments with heteronuclear NMR alleviated problems associated with large linewidths because heteronuclear couplings are large relative to the linewidths. (Note that the linewidths increase with increasing molecular weight because of slower tumbling of the molecule in solution; as a consequence, the efficiency of magnetization transfer through bonds employed in NMR experiments decreases.) Concomitant with the spectroscopic advances, improvements have been made with respect to the accuracy with which macromolecular structures can be determined. Thus it is now possible to determine the structures of proteins in the 15–35 kDa range at a resolution comparable to ~2.5 Å resolution crystal structures. The upper limit of applicability now is probably around 60–70 kDa, the largest single-chain proteins solved to date are ~30 kDa, comprising in excess of 240 residues [6••–8••], and the largest protein solved to date, the trimeric ectodomain of simian immunodeficiency virus (SIV) gp41, is ~44 kDa (Figure 1) [9••]. In addition, the structure of the elongation initiation factor 4E (a translation factor) solubilized by the detergent CHAPS, with an overall apparent molecular weight (protein plus CHAPS micelle) of ~40 kDa, has been determined [10••]. In this review we will summarize recent developments that have advanced the frontiers of applicability of NMR as a method of 3D structure determination of macromolecules in solution.

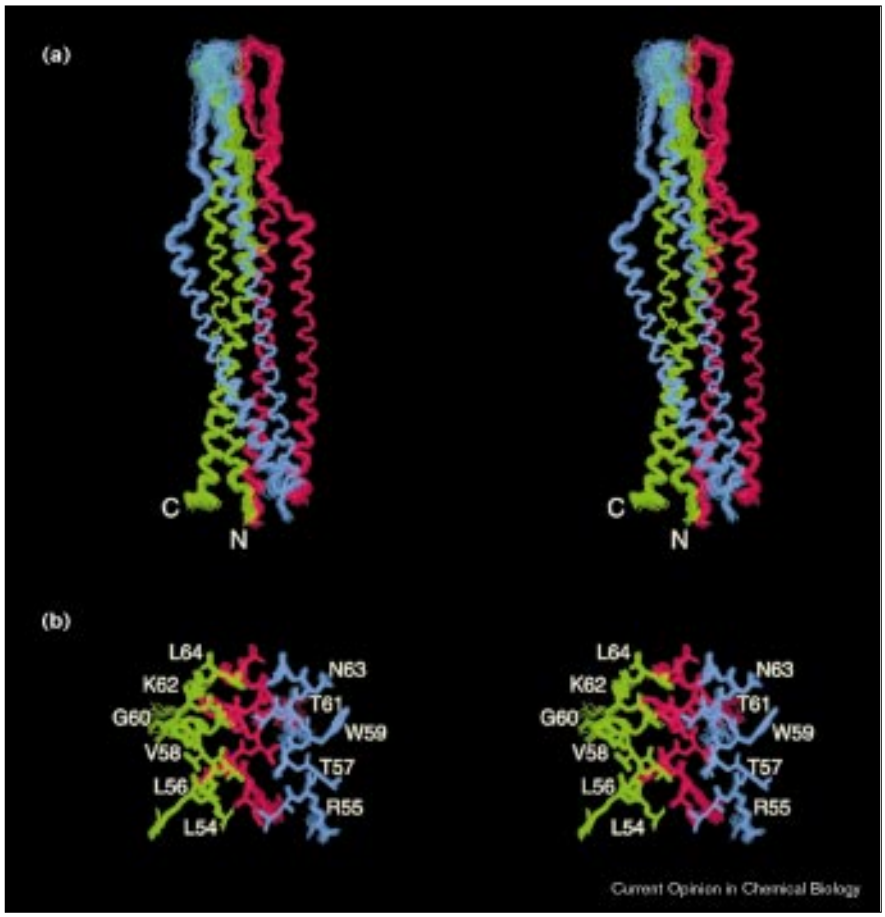
General strategy

The power of NMR over other spectroscopic techniques results from the fact that every NMR active nucleus gives rise to an individual resonance in the spectrum that can be resolved by higher dimensional (i.e. 2D, 3D or 4D) techniques. Bearing this in mind, the principles of structure determination by NMR can be summarized as follows: sequential resonance assignment using a combination of through-bond and through-space correlations; torsion angle determination and stereospecific assignments at chiral centers using three-bond scalar couplings combined, where appropriate, with intra-residue and sequential inter-residue nuclear Overhauser enhancement (NOE) data; identification of through-space connectivities between protons separated by less than 5 Å; and calculation of 3D structures on the basis of the experimental NMR restraints using one or more of a number of algorithms, such as distance geometry and/or simulated annealing [11••].

Sequential assignment

Conventional sequential resonance assignment, which has been applied to proteins up to about 100 residues, albeit with considerable effort, relies on 2D homonuclear ^{1}H–^{1}H through-bond correlation experiments to identify amino acid spin systems coupled with 2D ^{1}H–^{1}H NOE experiments to identify through-space (5 Å) sequential connectivities along

Figure 1



Stereoviews showing best-fit superpositions of **(a)** the backbone and **(b)** selected sidechains of the ensemble of 40 simulated annealing structures of the ectodomain of SIV gp41, derived from 2160 experimental NMR restraints per subunit, including 232 unambiguous intersubunit NOEs identified using various combinations of 1:1 mixtures of isotopically labeled (^{13}C, ^{15}N and ^{2}H) and unlabeled (^{12}C, ^{14}N and ^{1}H) subunits. Subunits A, B and C are displayed in blue, red and green, respectively. The location of the amino- and carboxyl termini of subunit B are indicated in (a).

the backbone of the type CαH(i)–NH(i+1,2,3,4), NH(i)–NH(i±2) and CαH(i)–CβH(i+3) [2,3]. (Cα represents mainchain carbons and Cβ represents sidechain carbons; i represents the reference amino acid residue, with i+1 indicating the next residue in the chain.)

As stated earlier, for larger proteins, the spectral complexity is such that 2D experiments no longer suffice, and it is essential to increase the spectral resolution by increasing the dimensionality of the spectra [1]. In some cases it is still possible to apply the same sequential assignment strategy by making use of 3D heteronuclear (^{15}N or ^{13}C)- separated experiments to increase the spectral resolution [1,12]. Frequently, however, numerous ambiguities still remain and it is advisable to adopt a sequential assignment strategy based solely on through-bond correlations involving well-defined heteronuclear one-bond ($^{1}J_{NC\alpha}$, $^{1}J_{NCO}$, $^{1}J_{C\alpha CO}$, $^{1}J_{CC}$) and two-bond ($^{2}J_{NC\alpha}$) scalar couplings along the polypeptide chain [1,13,14]. With the advent of pulsed-field gradients [15,16] it is now possible to acquire as few as two to four scans per increment without any loss in sensitivity (other than that due to the reduction in measurement time) such that each 3D experiment can be recorded in as little as seven hours. In most cases, however, signal-to-noise requirements necessitate one to three days measuring time, depending on the experiment.

For proteins greater than ~25 kDa, the assignment of the backbone and sidechain carbons is facilitated by making use of a sample in which the nonexchangeable (carbon-attached) protons are deuterated, resulting in a significant reduction in linewidths [6••,17–19,20••]. Narrower lines are the result of a substantial increase in the relaxation times of carbon and proton spins in proximity to substituted deuterons because of the approximately sixfold lower gyromagnetic ratio of deuterons relative to protons. Thus, for example, in the case of the 30 kDa amino-terminal domain of Enzyme I (a protein involved in bacterial sugar transport), the average transverse relaxation time (T_2) for the backbone amides is reduced from ~13 ms in the protonated sample to ~28 ms in the perdeuterated sample [6••].

In addition to perdeutration, it is also possible to significantly narrow the linewidths in a so-called 'transverse relaxation-optimized spectroscopy (TROSY) experiment', by suppressing transverse relaxation by the use of constructive intereference between dipole–dipole coupling and chemical shift anisotropy [21••]. The TROSY

sequence can be incorporated into numerous multidimensional experiments and its effectiveness increases with increasing field strength. For a 150 kDa protein at a spectrometer frequency of 750 MHz, it is predicted that the residual linewidths of the ^{15}N and NH resonances in a TROSY experiment will be ~10 Hz and ~45 Hz, respectively, in a protonated sample, and ~5 Hz and 15 Hz, respectively, in a perdeuterated sample. Thus, the TROSY sequence, in conjunction with appropriate levels of perdeuteration, promises to remove one of the major impediments in extending the NMR method to larger molecular weight proteins, namely broad linewidths arising from slow molecular tumbling.

Although perdeuteration permits the assignments of the carbon sidechains, it is still essential to assign the sidechain protons, since it is the through-space NOE interactions between protons that provide the main source of geometric information used to calculate the structures. For proteins up to 25 kDa, through-bond HCCH-correlated and total correlated spectroscopy experiments can be employed to transfer magnetization from protons to carbon, through neighbouring carbons, and back to protons. For proteins larger than 25 kDa, however, the sensitivity of these two experiments is markedly reduced as the ^{13}C linewidths approach the value of the $^{1}J_{CC}$ coupling (~30–35 Hz). An alternative, powerful, strategy is to rely on through-space correlations between adjacent carbon atoms (i.e. ^{13}C–^{13}C NOE) [22]. In contrast to through-bond correlations, the efficiency of transfer of the ^{13}C–^{13}C NOE increases with both increasing molecular weight and increasing field strength.

Torsion angles and sidechain rotamers

Torsion angle restraints can be readily derived from coupling constant data because simple geometric relationships exist between three-bond couplings and torsion angles. In simple systems, the coupling constant can be measured directly from the in-phase or antiphase splitting of a particular resonance in the 1D or 2D spectrum. For larger systems where the linewidths exceed the coupling, it becomes difficult to extract accurate couplings in this manner. An alternative approach involves the use of exclusive correlated spectroscopy (ECOSY) to generate reduced cross-peak multiplets [23]. While this permits accurate couplings to be obtained, the sensitivity of ECOSY experiments is generally quite low. Furthermore, in multidimensional experiments its utility is restricted by the fact that the couplings have to be measured in the indirectly detected frequency dimensions, and hence are influenced by limited digital resolution. To circumvent these limitations, a series of highly sensitive quantitative J correlation experiments have been developed [24–28]. These experiments quantitate the loss in magnetization when dephasing caused by coupling is active rather than inactive. In some quantitative J correlation experiments, the coupling is obtained from the ratio of cross-peak to diagonal-peak intensities. In others, it is obtained by the ratio of the cross-peaks obtained in two separate experiments (with the coupling active and inactive), recorded in an interleaved manner. Particularly useful couplings are $^{3}J_{HN\alpha}$ and $^{3}J_{C'C'}$, which are related to the backbone torsion angle ϕ, $^{3}J_{C\gamma C'}$, $^{3}J_{C\gamma N}$, $^{3}J_{NH\beta}$, $^{3}J_{C'C\beta}$ and $^{3}J_{H\alpha H\beta}$, which are related to the $\chi 1$ sidechain torsion angle, and $^{3}J_{CC}$ and $^{3}J_{CH}$, which are related to the χ_2 and χ_3 sidechain torsion angles of leucine, isoleucine and methionine. Similar quantitative J correlation experiments can also be applied to nucleic acids [28].

For smaller proteins it is often possible to obtain stereospecific assignments of β-methylene protons on the basis of a qualitative interpretation of the homonuclear $^{3}J_{\alpha\beta}$ coupling constants and the intra-residue NOE data involving the NH, CαH and CβH protons [2,3]. A more rigorous approach, which also permits one to obtain ϕ, ψ backbone and χ_1 sidechain torsion angle restraints, involves the application of a conformational grid search of ϕ,ψ,χ_1 space on the basis of the homonuclear $^{3}J_{HN\alpha}$ and $^{3}J_{\alpha\beta}$ coupling constants (which are related to ϕ and χ_1, respectively), and the intra-residue and sequential inter-residue NOEs involving the NH, CαH and CβH protons [29,30]. This information can be supplemented and often supplanted by the measurement of heteronuclear couplings by quantitative J correlation spectroscopy. For larger proteins, the most useful couplings in this regard are the $^{3}J_{C\gamma CO}$ and $^{3}J_{NC\gamma}$ couplings, involving the aromatic, methyl and methylene Cγ atoms, which are sufficient, when used in combination, to derive the appropriate χ_1 sidechain rotamer.

Very recently, a new method, which does not rely on coupling constants, has been developed to directly measure angles between bond vectors [31••,32••]. It is based on cross-correlated relaxation, and to date has been use to define the backbone torsion angle ψ, either by measuring dipole–dipole cross-correlated relaxation of double-quantum and zero-quantum coherences involving the inter-residue dipolar fields of the ^{15}N–H(i) and ^{13}Cα–Hα(i-1) bond vectors [31••], or by measuring the cross-correlated relaxation between the ^{13}Cα–Hα dipolar interaction and the ^{13}C′(carbonyl) chemical shift anisotropy [32••].

Assignment of through-space proton–proton interactions within a protein

While the panoply of 3D heteronuclear experiments is sufficient for the purposes of spectral assignment, yet further increases in resolution because of spectral overlap are required for the reliable identification of NOE through-space interactions. This can be achieved by extending the dimensionality still further to four dimensions [1]. In this manner, each ^{1}H–^{1}H NOE interaction is specified by four chemical shift coordinates, the two protons giving rise to the NOE and the heavy atoms to which they are attached.

Because the number of NOE interactions present in each 2D plane of a 4D ^{13}C/^{15}N, ^{13}C/^{13}C or ^{15}N/^{15}N-separated NOE spectrum is so small, the inherent resolution in a

4D spectrum is extremely high, despite the low level of digitization [1]. Indeed, one can expect to obtain both good sensitivity and resolution for proteins up to 400 residues. Thus, once complete ^{1}H, ^{15}N and ^{13}C assignments are obtained, analysis of 4D $^{15}N/^{13}C$, $^{13}C/^{13}C$ and $^{15}N/^{15}N$-separated NOE spectra should permit the assignment of numerous NOE interactions in a relatively straightforward manner [1]. The first successful application of these methods to the structure determination of a protein greater than 15 kDa was achieved in 1991, with the determination of the solution structure of interleukin-1β, a protein of 18 kDa and 153 residues [33]. This has now been extended to three ~30 kDa single chain proteins: the amino-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate : sugar phosphotransferase system (259 residues) [6••], the serine protease PB92 (269 residues) [7••] and the methyltransferase ermAm (245 residues) [8••], as well as to the trimeric 44 kDa ectodomain of SIV gp41 [9••].

For proteins larger than 30 kDa, it may not always be possible to obtain complete sidechain proton assignments either because of spectral overlap or the absence of through-bond connectivities because of large linewidths. In such cases, appropriate use of deuteration can be extremely helpful. One approach is to record a 4D $^{15}N/^{15}N$-separated NOE spectrum on a perdeuterated protein in which only the amides are protonated [17,18]. This yields exceptional sensitivity because many spin-diffusion pathways are removed, thereby permitting long mixing times to be employed. More recently, labeling strategies have been developed in which the amides and methyl groups are protonated, while all other groups are deuterated [19,20••]. This labeling profile retains the advantage of deuteration in triple resonance assignment experiments while ensuring the presence of a reasonable number of protons at useful locations for the purpose of structure determination. Indeed, model calculations have shown that it is possible to obtain low resolution global folds on the basis of NH–NH, NH–methyl and methyl–methyl distances [19,20••].

Protein–ligand and protein–protein complexes

Providing one of the partners in a complex (e.g. a peptide, an oligonucleotide, a drug, and so on) presents a relatively simple spectrum that can be assigned by 2D methods, the most convenient strategy for dealing with protein complexes involves one in which the protein is labeled with ^{15}N and ^{13}C and the partner is unlabeled (i.e. at natural isotopic abundance). It is then possible to use a combination of heteronuclear filtering and editing to design experiments in which correlations involving only protein resonances, only ligand resonances, or only through-space interactions between ligand and protein are observed (see [34] for a recent review). These experiments have been successfully employed in a number of laboratories for a range of systems, including protein–drug, protein–peptide and protein–nucleic acid complexes.

Oligomeric proteins represent complexes between identical subunits. For multimeric proteins, additional labeling schemes can also be used to facilitate the identification of intermolecular NOEs. For example, 1:1 mixtures of $^{13}C/^{15}N/^{1}H$: $^{12}C/^{14}N/^{1}H$, $^{13}C/^{14}N/^{1}H$: $^{12}C/^{15}N/^{2}H$ and $^{13}C/^{15}N/^{2}H$: $^{12}C/^{14}N/^{1}H$ enable one to record high sensitivity 2D, 3D and 4D experiments to specifically observe NOEs from protons attached to ^{13}C or ^{15}N to protons attached to ^{12}C or ^{14}N, from protons attached to ^{13}C to protons attached to ^{15}N, and from protons attached to ^{15}N to protons attached to ^{12}C or ^{14}N, respectively [34]. This has recently been applied to the trimeric 44 kDa ectodomain of SIV gp41 [9••].

Additional methods of structure refinement

The NOE-derived interproton distance and torsion angle restraints that are traditionally employed in NMR structure determination can be supplemented by direct refinement against a number of other NMR observables in a relatively straightforward manner (see [11••] for a recent review). These include three-bond coupling constants (related to torsion angles), three-bond amide deuterium isotope effects on $^{13}C\alpha$ shifts (related to the backbone ψ angle [35•]), ^{13}C secondary chemical shifts (related to the backbone ϕ and ψ angles), and ^{1}H chemical shifts (which are influenced by short-range ring current effects from aromatic groups, magnetic anisotropy of C=O and C-N bonds and electric field effects arising from charged groups). The rationale behind including these restraints is twofold: they are easily measured and therefore represent a useful source of additional structural restraints, and it is generally found that the agreement between observed and calculated values for these various parameters is better for high-resolution X-ray structures than for the corresponding high-resolution NMR structures refined in the absence of these restraints. Inclusion of these restraints has little impact on precision but does increase the accuracy of the structures.

Further improvements in the quality of structures generated from NMR data can be obtained by using a conformational database potential derived from dihedral angle relationships in databases of high-resolution highly refined protein and nucleic acid crystal structures [36,37]. The rationale for this procedure is based on the observation that uncertainties in the description of the nonbonded contacts present a key limiting factor in the attainable accuracy of protein NMR structures and that the nonbonded interaction terms presently used have poor discriminatory power between high and low probability local conformations. The idea behind the conformational database potential is to bias sampling during simulated annealing refinement to conformations that are likely to be energetically possible by effectively limiting the choices of dihedral angles to those that are known to be physically realizable. In this manner, the variability in the structures is primarily a function of the experimental restraints, rather than an artifact of a poor nonbonded interaction model. This can be readily achieved

without compromising the agreement with the experimental restraints and the deviations from idealized covalent geometry that remain within experimental error.

Long-range structural restraints

Until very recently, structure determination by NMR has relied exclusively on restraints whose information is entirely local and restricted to atoms close in space, specifically NOE-derived short (<5 Å) interproton distance restraints, supplemented by coupling constants, ^{13}C secondary shifts and ^{1}H shifts. The success of the NMR method is because of the fact that short interproton distances between units far apart in a linear array are conformationally highly restrictive [1]; however, there are numerous cases where restraints that define long-range order would be highly desirable. In particular, they would permit the relative positioning of structural elements that do not have many short interproton distance contacts between them. Examples of such situations include modular and multidomain proteins and linear nucleic acids. Two novel approaches have recently been introduced that directly provide restraints that characterize long-range order *a priori* [38••,39••]. The first relies on the dependence of heteronuclear (^{15}N or ^{13}C) longitudinal (T_1) relaxation times and T_2 (specifically T_1:T_2 ratios) on rotational diffusion anisotropy [38••], and the second relies on residual dipolar couplings in oriented macromolecules [39••,40•]. The two methods provide restraints that are related in a simple geometric manner to the orientation of internuclear vectors relative to an external axis system (represented, for example, by the diffusion, magnetic susceptibility or molecular alignment tensors).

For the heteronuclear ^{15}N T_1/T_2 method to be applicable the molecule must tumble anisotropically (i.e. it must be nonspherical). The minimum ratio of the diffusion anisotropy for which heteronuclear T_1/T_2 refinement will be useful depends entirely on the accuracy and uncertainties in the measured T_1:T_2 ratios. In practice, the difference between the maximum and minimum observed T_1:T_2 ratios must exceed the uncertainty in the measured T_1:T_2 ratios by an order of magnitude. This typically means that the diffusion anisotropy should be greater than ~1.5 [38••].

Likewise, the applicability of the residual dipolar coupling method depends on the magnitude of the the degree of alignment of the molecule in the magnetic field. In practice, the residual dipolar couplings must exceed the uncertainty in their measured values by an order of magnitude. For molecules in isotropic solution this typically means that the magnetic susceptibility anisotropy should be ~ 20×10^{-34} m^3/molecule, which is about 20-fold greater than the magnetic susceptibility anisotrophy for benzene [39••]. Magnetic susceptibility anisotropies of this magnitude are only obtained for a limited number of systems (e.g. paramagnetic proteins, nucleic acids and proteins complexed to nucleic acids). Recently it has been shown that moderate degrees of alignment, while retaining the spectral resolution, sensitivity and simplicity obtained in isotropic aqueous medium, can be obtained by dissolving macromolecules in a dilute liquid crystalline phase [41••,42••] of so-called 'bicelles' [43], or rod-shaped virus particles such as the filamentous phage fd or tobacco mosaic virus (GM Clove, MA Starich, AM Gronenborn, unpublished data; see Note added in proof). By these means, accurate measurement of residual dipolar couplings for a variety of different fixed distance internuclear vector types, including one-bond ^{15}N–$^{1}H_N$, ^{13}C–^{1}H, $^{13}C\alpha$–$^{13}C'$ and ^{15}N–$^{13}C'$ dipolar couplings, as well as two-bond $^{1}H_N$–13C′, can be readily obtained.

Refinement against ^{15}N–^{1}H and $^{13}C\alpha$–H dipolar couplings induced by the magnetic field has been successfuly applied to a number of protein–DNA complexes [39••,44••,45]. Refinement against a much more extensive set of dipolar couplings measured using a dilute liquid crystal phase of bicelles has recently been carried out for the protein cyanovirin-N, a potent HIV-inactivating protein [46••] and the human barrier-to-autointergration factor BAF (M Cai *et al.*, unpublished data; see Note added in proof).

A key aspect of the use of either ^{15}N T_1:T_2 ratios or dipolar couplings for structure refinement is the determination of the magnitude of the anisotropy and rhombicity in the absence of any structural information. These quantities can be obtained by examining the distribution of the measured ^{15}N T_1:T_2 ratios [47••] or dipolar couplings [48•] that have a powder pattern-like appearance in the case of ^{15}N T_1:T_2 ratios and an exact powder pattern appearance in the case of dipolar coupling measurements.

Conclusions and future perspectives

The recent development of a whole range of highly sensitive multidimensional heteronuclear edited and filtered NMR experiments has propelled the field of protein structure determination by NMR into larger molecular weight ranges. Proteins and protein complexes in the 20–50 kDa range are now amenable to detailed structural analysis in solution. In addition, although not touched upon in this review, NMR offers a unique means of probing molecular motions on the picosecond to nanosecond and on the microsecond to millisecond time scales (see [49••] for a recent review).

Despite these advances, it should always be borne in mind that there are a number of key requirements that have to be satisfied to permit a successful structure determination of larger proteins and protein complexes by NMR. The protein in hand must be soluble and should not aggregate up to concentrations of about 0.5–1 mM; it must be stable at room temperature (or slightly higher) for considerable periods of time (particularly as it may take several months of measurement time to acquire all the necessary NMR data); it should not exhibit significant conformational heterogeneity that could result in extensive line broadening; and finally it must be amenable to uniform ^{15}N and ^{13}C labeling. At the present time there are still only relatively few examples in the literature of proteins in the 15–30 kDa range that have been

solved by NMR. Likewise, only a handful of protein–DNA and protein–peptide complexes and oligomers have been determined to date using these methods. One can anticipate, however, that over the next few years, by the widespread use of multidimensional heteronuclear NMR experiments, coupled with semiautomated assignment procedures, many more NMR structures of proteins and protein complexes will become available.

Note added in proof

The papers referred to in the text as (GM Clore, MA Starich, AM Gronenborn, unpublished data) and (M Cai *et al*., unpublished data) have now been accepted for publication [50••,51••].

Acknowledgements

We thank A Bax and DS Garrett for numerous stimulating discussions. The work in the authors' laboratory was in part supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Clore GM, Gronenborn AM: **Structures of larger proteins: three- and four-dimensional heteronuclear NMR spectroscopy.** *Science* 1991, **252**:1390-1399.

2. Wüthrich K: *NMR of Proteins and Nucleic Acids*. New York: Wiley; 1996.

3. Clore GM, Gronenborn AM: **Determination of three-dimensional structures of proteins in solution by nuclear magnetic resonance spectroscopy.** *Protein Eng* 1987, **1**:275-288.

4. Dyson HJ, Gippert, GP, Case DA, Holmgren A, Wright PE: **Three-dimensional solution structure of the reduced form of *Escherichia coli* thioredoxin determined by nuclear magnetic resonance spectroscopy.** *Biochemistry* 1990, **29**:4129-4136.

5. Forman-Kay JD, Clore GM, Wingfield PT, Gronenborn AM: **The high resolution three-dimensional structure of reduced recombinant human thioredoxin in solution.** *Biochemistry* 1991, **30**:2685-2698.

6. •• Garrett DS, Seok YJ, Liao DI, Peterkofksy A, Gronenborn AM, Clore GM: **Solution structure of the 30 kDa N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR.** *Biochemistry* 1997, **36**:2517-2530.
This paper presents the structure determination of one of the first single chain 30 kDa proteins solved by NMR and describes in detail how this was accomplished, including the use of perdeuteration.

7. •• Martin JR, Mulder FAA, Karimi-Nejad Y, van der Zwan J, Mariani M, Schipper D, Boelens R: **The solution structure of serine protease PB92 from *Bacillus alcalophilus* presents a rigid fold with a flexible substrate-binding site.** *Structure* 1997, **5**:521-532.
This presents the structure determination of one of the first single chain 30 kDa proteins by NMR.

8. •• Yu L, Petros AM, Schnuchel A, Zhong P, Severin JM, Walker K, Holzman TF, Fesik SW: **Solution structure of an rRNA methyltransferase (ErmAm) that confers MLS antibiotic resistance.** *Nat Struct Biol* 1997, **4**:483-489.
This presents the structure determination of one of the first single chain 30 kDa proteins by NMR.

9. •• Caffrey M, Cai M, Kaufman J, Stahl SJ, Wingfield PT, Covell DG, Gronenborn AM, Clore GM: **Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41.** *EMBO J* 1998, **17**:4572-4584.
This paper describes the structure determination of a symmetric 44 kDa trimer. The trimeric nature of the structure makes this a particularly difficult task for NMR, and various NMR experiments for observing intermolecular contacts in proteins of this size are presented.

10. •• Matsuo H, Li HJ, McGuire AM, Fletcher CM, Gingras AC, Sonenberg N, Wagner G: **Structure of translation factor EIF4E bound to m7GDP and interaction with 4E-binding protein.** *Nat Struct Biol* 1997, **4**:717-724.
This paper describes the use of CHAPS to increase the solubility of a protein and shows that the structure of a protein–micelle complex of ~40 kDa can be solved by NMR.

11. •• Clore GM, Gronenborn AM: **New methods of structure refinement for macromolecular structure determination by NMR.** *Proc Natl Acad Sci USA* 1998, **95**:5891-5898.
This paper summarizes a variety of new refinement methods for NMR structure determination, including the use of coupling constants, ^{13}C shifts, ^{1}H shifts, T_1:T_2 ratios, dipolar couplings and conformational databases.

12. Fesik SW, Zuiderweg ERP: **Heteronuclear three-dimensional NMR spectroscopy of isotopically labeled biological macromolecules.** *Q Rev Biophys* 1990, **23**:97-131.

13. Clore GM, Gronenborn AM: **Applications of three- and four-dimensional heteronuclear NMR spectroscopy to protein structure determination.** *Prog NMR Spectrosc* 1991, **23**:43-92.

14. Bax A, Grzesiek S: **Methodological advances in protein NMR.** *Accounts Chem Res* 1993: **26**:131-138.

15. Bax A, Pochapsky SS: **Optimized recording of heteronuclear multidimensional NMR spectra using pulsed field gradients.** *J Magn Reson* 1992, **99**:638-643.

16. Kay LE, Keifer P, Saarinen T: **Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity.** *J Am Chem Soc* 1992, **114**:10663-10665.

17. Grzesiek S, Wingfield PT, Stahl, SJ, Bax A: **Four-dimensional ^{15}N-separated NOESY of slowly tumbling perdeuterated ^{15}N-enriched proteins: application ot HIV-1 nef.** *J Am Chem Soc* 1995, **117**:9594-9595.

18. Venters RA., Metzler WJ, Spicer LD, Mueller L, Farmer BT: **Use of $^{1}H_N$-$^{1}H_N$ NOEs to determine protein global folds in perdeuterated proteins.** *J Am Chem Soc* 1995, **117**:9592-9593.

19. Metzler WJ, Wittekind M, Goldfarb V, Mueller L, Farmer BT: **Incorporation of $^{1}H/^{13}C/^{15}N$-{ile, leu, val} into a perdeuterated, ^{15}N-labeled protein: potential in structure determination of larger proteins by NMR.** *J Am Chem Soc* 1996, **118**:6800-6801.

20. •• Gardner KH, Rosen MK, Kay LE: **Global folds of highly deuterated, methyl protonated proteins by multidimensional NMR.** *Biochemistry* 1997, **36**:1389-1401.
This paper presents a detailed study of the sort of structures that can be obtained using restraints derived from proteins that are fully perdeuterated with the exception of the amide and methyl groups.

21. •• Pervushin K, Riek R, Wider G, Wüthrich K: **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* 1997, **94**:12366-12371.
This paper describes a simple method, based on the use of constructive intereference between dipole–dipole coupling and chemical shift anisotropy, for eliminating in part one of the major obstacles, namely line broadening due to slow tumbling, for studying large proteins by NMR

22. Fischer MWF, Leng L, Zuiderweg ERP: **Use of ^{13}C-^{13}C NOE for assignment of NMR lines of larger labeled proteins at larger magnetic fields.** *J Am Chem Soc* 1996, **118**:12457-12458.

23. Griesinger C, Sørensen OW, Ernst RR: **Correlation of connected transitions by two-dimensional NMR spectroscopy.** *J Chem Phys* 1986, **85**:6837-6852.

24. Bax A, Vuister GW, Grzesiek S, Delaglio F, Wang AC. Tschudin R, Zhu G: **Measurement of homo- and heteronuclear J couplings from quantitative J correlation.** *Methods Enzymol* 1994, **239**:79-105.

25. Ju JS, Bax A: **Measurement of three-bond ^{13}C-^{13}C J couplings between carbonyl and carbonyl/carboxyl carbons in isotopically enriched proteins.** *J Am Chem Soc* 1996, **118**:8170-8171.

26. Hu JS, Bax A: **χ_1 angle information from a simple two-dimensional NMR experiment that identified trans $^{3}J_{NC\gamma}$ coupling in isotopically enriched proteins.** *J Biomol NMR* 1997, **9**:323-328.

27. Hu JS, Grzesiek S, Bax A: **Two dimensional NMR methods for determining χ_1 angles of aromatic residues from three-bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ couplings.** *J Am Chem Soc* 1997, **119**:1803-1804.

28. Clore GM, Murphy EC, Gronenborn AM, Bax A: **Determination of three-bond $^{1}H3'$-^{31}P couplings in nucleic acids and protein-nucleic acid complexes by quantitative J correlation spectroscopy.** *J Magn Reson* 1998, in press.

29. Güntert P, Braun W, Billeter M, Wüthrich K: **Automated stereospecific ^{1}H assignments and their impact on the precision of protein structure determinations in solution.** *J Am Chem Soc* 1989, **111**:3397-4004.

30. Nilges M, Clore GM, Gronenborn AM: **^{1}H-NMR stereospecific assignments by conformational database searches.** *Biopolymers* 1990, **29**:813-822.

31. •• Reif B, Hennig M, Griesinger C: **Direct measurement of angles between bond vectors in high-resolution NMR.** *Science* 1997, **276**:1230-1233.
This presents a new method for directly determining angles between bond vectors by measuring the cross-correlated relaxation between two dipolar fields, and applies it to the determination of the backbone torsion angle ψ.

32. •• Yang D, Konrat R, Kay LE: **A multidimensional NMR experiment for measurement of the protein dihedral angle ψ based on cross-correlated relaxation between $^{1}H\alpha$-$^{13}C\alpha$ dipolar and $^{13}C'$(carbonyl) chemical shift anisotropy mechanisms.** *J Am Chem Soc* 1997, **119**:11938-11940.
This paper presents a new method for determining angles between bond vectors by measuring the cross-correlated relaxation between dipolar and chemical shift anisotropy relaxation mechanisms, and applies it to the determination of the backbone torsion angle ψ.

33. Clore GM, Wingfield PT, Gronenborn AM: **High resolution three-dimensional structure of interleukin-1β in solution by three and four dimensional nuclear magnetic resonance spectroscopy.** *Biochemistry* 1991, **30**:2315-2323.

34. Clore GM, Gronenborn AM: **Determining the structures of large proteins and protein complexes by NMR.** *Trends Biotechnol* 1998, **16**:22-34.

35. • Ottiger M, Bax A: **An empirical correlation between amide deuterium isotope effects on $^{13}C\alpha$ chemical shifts and protein backbone conformation.** *J Am Chem Soc* 1997, **119**:8070-8075.
This paper describes the existence of a simple Karplus-type relationship between the three-bond amide deuterium isotope effect on $^{13}C\alpha$ shifts and the ψ backbone torsion angle.

36. Kuszewski J, Gronenborn AM, Clore GM: **Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases.** *Protein Sci* 1996, **5**:1067-1080.

37. Kuszewski J, Gronenborn AM, Clore GM: **Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids.** *J Magn Reson* 1997, **125**:171-177.

38. •• Tjandra N, Garrett DS, Gronenborn AM, Bax A, Clore GM: Defining long range order in **NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy.** *Nat Struct Biol* 1997, **4**:443-449.
This paper describes how the dependence of heteronuclear relaxation in the form of ^{15}N T_1:T_2 ratios can be used as restraints that define long-range order *a priori*, and applies it to the two domain, 30 kDa protein enzyme I.

39. •• Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A: **Use of dipolar ^{1}H-^{15}N and ^{1}H-^{13}C couplings in structure determination of magnetically oriented macromolecules in solution.** *Nat Struct Biol* 1997, **4**:732-738.
This paper describes how dipolar couplings arising from magnetic susceptibility anisotropy can be used to provide restraints that define long-range order *a priori*, and applies it to a complex of the transcription factor GATA-1 with DNA. It is shown that incorporation of N–H and Cα–H dipolar couplings results in a significant increase in the percentage of residues in the most favourable region of the Ramachandran ϕ,ψ map.

40. • Clore GM, Gronenborn AM, Tjandra N: **Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude.** *J Magn Reson* 1998, **131**:159-162.
This paper provides details of how to refine structures against dipolar couplings.

41. •• Tjandra N, Bax A: **Direct measurement of distances and angles in biomolecules by NMR in dilute liquid crystalline medium.** *Science* 1997, **278**:1111-1114.
This paper describes how moderate degrees of alignment while retaining the resolution, sensitivity and simplicity of the isotropic NMR spectrum, can be obtained by the use of a dilute liquid crystalline medium of bicelles, thereby permitting the measurement of a wide range of dipolar couplings.

42. •• Ottiger M, Bax A: **Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings of macromolecules.** *J Biomol NMR* 1998, in press.
This paper describes in detail how to prepare bicelles for orienting macromolecules and characterizes the ranges of concentration, temperature, pH and salt over which the liquid crystalline phase is stable.

43. Sanders CR, Schwonek JP: **Characterization of magnetically orientable bilayers in mixtures of dihexanoylphophatidylcholine and dimyristoylphophatidylcholine by solid-state NMR.** *Biochemistry* 1992, **31**:8898-8905.

44. •• Starich MR, Wikström M, Arst HN, Clore GM, Gronenborn AM: **The solution structure of a fungal AREA protein-DNA complex reveals an alternative binding mode for the basic carboxyl tail of GATA factors.** *J Mol Biol* 1998, **277**:605-620.
Refinement against dipolar couplings is shown to play a key role in the ability to determine the conformation of the carboxyl-terminal tail, which binds in the minor groove of the DNA.

45. Starich MR, Wikström M, Schumacher S, Arst HN, Gronenborn AM, Clore GM: **The solution structure of the leu22→val mutant AREA DNA binding domain complexed with a TGATAG core element defines a role for hydrophobic packing in the determination of specificity.** *J Mol Biol* 1998, **277**:621-634.

46. •• Bewley CA, Gustafson KR, Boyd MR, Covell DG, Bax A, Clore GM, Gronenborn AM: **Solution structure of cyanovirin-N, a potent HIV-inactivating protein from the cyanobacterium *Nostoc ellipsosporum*.** *Nat Struct Biol* 1998, **5**:571-578.
This paper presents the first example of a structure refined against five different dipolar couplings, namely the one-bond N–H, Cα–H, N–C′ and Cα–C′ dipolar couplings and the two-bond H_N–C′ dipolar coupling. It is shown that this increases the precision of the resulting structures.

47. •• Clore GM, Gronenborn AM, Szabo A, Tjandra N: **Determining the magnitude of the fully asymmetric diffusion tensor from heteronuclear relaxation data in the absence of structural information.** *J Am Chem Soc* 1998, **120**:4889-4890.
It is shown that the magnitude of the fully asymmetric diffusion tensor can be readily obtained in the absence of any structural information or resonance assignments by examining the distribution of measured ^{15}N T_1:T_2 values, which has a powder pattern-like appearance. This not only permits the use of ^{15}N T_1:T_2 ratios for refinement, but also enables one to derive information on the hydrodynamic properties and shape of the macromolecule under consideration.

48. • Clore GM, Gronenborn AM, Bax A: **A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information.** *J Magn Reson* 1998, **133**:216-221.
It is shown that the magnitude of the fully asymmetric alignment tensor can be obtained in the absence of any structural information by examining the distribution of measured dipolar couplings, which has a powder pattern appearance. It is demonstrated that the accuracy with which the magnitude of the tensor can be obtained is improved by making use of several different dipolar couplings and normalizing them, thereby ensuring a more uniform distribution of vectors.

49. •• Palmer AG: **Probing molecular motion by NMR.** *Curr Opin Struct Biol* 1997, **7**:732-737.
This is a thorough review of recent advances in the study of dynamics by NMR.

50. •• Clove GM, Starich MA, Gronenborn AM: **Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses.** *J Am Chem Soc* 1998, in press.
This paper describes how moderate degrees of alignment, permitting the successful measurement of residual dipolar couplings, can readily be obtained using colloidal suspensions of filamentous phage such as fd or tobacco mosaic virus. It is also shown that virus particles can be used successfully in cases where bicelles are unstable (either becuase of interaction with the protein or because of solution conditions such as pH and temperature), and that the nematic phase is maintained over a wide range of pH and temperature.

51. •• Cai M, Yuang Y, Zheng R, Wei SQ, Ghirlando R, Lee MS, Craigie R, Gronenborn AM, Clore GM: **Solution structure of the cellular factor of BAF responsible for protecting retroviral DNA from autointegration.** *Nat Struct Biol* 1998, in press.
This paper describes the first example of the use of multiple dipolar coupling refinement for the structure determination of a dimeric protein. In such cases, the dipolar couplings can uniquely define the relative orientation of the subunits.