Theoretical and computational advances in biomolecular NMR spectroscopy

G Marius Clore* and Charles D Schwieters†

Recent developments in experimental and computational aspects of NMR spectroscopy have had a significant impact on the accuracy and speed of macromolecular structure determination in solution, particularly with regard to systems of high complexity (such as protein complexes). These include experiments designed to provide long-range orientational and translational restraints, improvements in internal coordinate dynamics used for simulated annealing, and the development of database potentials of mean force to improve the description of the non-bonded contacts.

Addresses
*Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0510, USA; e-mail: clore@speck.niddk.nih.gov
†Computational Bioscience and Engineering Laboratory, Building 12A, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892-5624, USA; e-mail: charles.schwieters@nih.gov

Abbreviations
CSA   chemical shift anisotropy
FRET   fluorescence resonance energy transfer
IVM   internal variable dynamics module
NOE   nuclear Overhauser enhancement

Introduction
NMR is a powerful spectroscopic technique that permits the detailed study of the structure and dynamics of macro-molecules in solution [1,2]. In this brief review, we will consider various recent theoretical and computational developments in NMR that relate specifically to aspects of 3D structure determination.

The general problem posed in 3D structure determination by NMR involves finding the global minimum region of a target function (E_{tot}) comprising three terms that describe experimental restraints (E_{Expt}), covalent geometry (E_{Cov}) and non-bonded interactions ($E_{Non-bonded}$) [3]. The energy surface of this target function is rough and there are numerous local minima on the path to the global minimum. The global minimum region comprises a set of structures that are consistent with the experimental restraints. Probably the most powerful, and certainly the most widely employed, computational method used to locate the global minimum region of this target function involves the applica-tion of simulated annealing by means of molecular dynamics calculations either in Cartesian or torsion-angle space [3].

Despite recent developments relating to long-range orientational restraints [4••,5–10], the principal source of

geometric information used in any NMR structure determination still consists of short (<6 Å) interproton distance restraints derived from nuclear Overhauser enhancement (NOE) measurements [1–3]. The NOE data can be supplemented by a number of other restraints that are dependent on the close spatial proximity of atoms. These include torsion-angle restraints, which are usually derived from three-bond scalar couplings [11], but are also accessible from cross-correlation effects [12] and deuterium isotope shifts [13]; secondary carbon chemical shift restraints, which can be empirically related to backbone torsion angles in proteins; and ^{1}H chemical shift restraints, which are affected by ring current shifts arising from close proximity to aromatic rings, by electronic field effects arising from charged groups and by magnetic anisotropy arising from C′–O and C′–N bonds [3]. Longer range distance information (potentially up to ~25–30 Å) can also be obtained using approaches that entail paramagnetic relaxation effects [14••,15••,16]. In the general case, where a paramagnetic group is not an intrinsic component of the molecule of interest (e.g. a heme group or a metal center), this involves chemical modification with an appropriate paramagnetic label. Finally, long-range orientational information is accessible through the measurement of residual dipolar couplings [7–10], chemical shift anisotropy (CSA) [17,18•,19] and heteronuclear relaxation parameters [20].

The experimental data available from NMR are not sufficient in their own right to determine a 3D structure of a macromolecule *a priori*. Thus, the experimental data must be supplemented by covalent geometry restraints (i.e. bond lengths, angles, planarity and chirality), which can be assumed to be known to high accuracy from small molecule work, and by a force field describing the non-bonded contacts. The latter can have a major influence on the accuracy of the calculated structures and, in particular, on internal packing [21,22••]. In its simplest representation, the non-bonded potential term can be described by a simple repulsive form to prevent atomic overlap [3]. Recent progress in this area has been aimed at supplementing conventional descriptions of the non-bonded contacts by database potentials of mean force derived from a statistical analysis of high-resolution crystal structures [22••,23,24].

Many aspects of the computational approaches employed in NMR structure determination have already been reviewed in detail [3]. In this review, we will therefore summarize some recent progress related specifically to the use of long-range structural restraints (distance and orientational), improvements in computational aspects of simulated annealing, new strategies for rapidly determining interdomain or intersubunit orientations, new approaches

for rapid protein fold determination and the development of new database potentials of mean force to enhance the description of the non-bonded interactions.

Long-range orientational restraints

Long-range orientational restraints can be derived from the measurement of residual dipolar couplings [4••,5–10] and CSA [17,18•,19] in liquid crystalline media and, in suitable cases, from heteronuclear T_1/T_2 data [20]. The characteristic feature of these various parameters is that they yield direct geometric information on the orientation of an interatomic vector(s) with respect to an external axis system (e.g. the alignment tensor in liquid crystalline media, the diffusion tensor for relaxation measurements). The orientation of the bond vector is expressed in terms of two angles: θ, the angle between the interatomic vector and the z axis of the tensor; and ϕ, the angle that describes the position of the projection of the interatomic vector on the x-y plane of the tensor. The latter angle is only relevant when the system deviates significantly from axial symmetry. In structure refinement, it is convenient to represent the orientation of the tensor by an external axis system comprising four atoms: O, X, Y and Z, with O at the origin and the O-X, O-Y and O-Z vectors mutually perpendicular to each other [25]. The orientation of this axis system is allowed to float relative to that of the associated molecule(s). In the case of ^1H-^1H dipolar couplings, the measured dipolar coupling is dependent not only on the orientation of the interproton vector relative to the alignment tensor but also on the inverse cube of the distance (r) between the two protons, yielding translational as well as orientational information [26•].

For dipolar couplings, two additional unknowns need to be determined in order to make use of the data in the above representation: namely, the magnitudes of the axial and rhombic components of the tensor. These can usually be obtained from the distribution of measured dipolar couplings either by simple inspection [27] or, if the data are not extensive enough, by means of maximum likelihood analysis [28•]. In addition, the values of the axial and rhombic components of the tensor can be optimized either by a systematic grid search [25] or by treating them as floating variables during the course of the refinement itself [29,30•]. The latter may prove particularly valuable when the number and distribution of the measured dipolar couplings are limited.

In the case of CSA, the difference in chemical shift in isotropic and aligned media provides direct information on the orientation of the CSA tensor of a particular atom with respect to the alignment tensor [17]. In effect, this translates to restraints relating to the orientation of pairs of bond vectors relative to the alignment tensor. For example, in the case of ^{31}P CSA of the phosphate groups of DNA, the two vectors comprise the P–O1 and P–O2 bonds [18•]; in the case of the protein backbone carbonyl atom, the two vectors comprise the C′–O and C′(i)–N(i+1) bonds [19].

To make use of CSA, the magnitude of the alignment tensor needs to be determined from dipolar coupling measurements and the values of the principal components of the traceless CSA tensor for a given atom type have to be known *a priori*. The latter can be determined either from solid-state measurements or from solution studies in conjunction with a high-resolution crystal structure. Detailed solution measurements on ubiquitin (for which a high-resolution crystal structure is available) have been used to obtain accurate average values for the principal components of the backbone amide proton, ^{15}N and $^{13}C′$ CSA tensors in proteins [17]. The principal components of the ^{31}P CSA tensor in nucleic acids have been determined from solid-state measurements [31]. The potential utility of CSA restraints has recently been demonstrated for both a short duplex DNA [18•] and a multidomain protein [19].

In the case of heteronuclear T_1/T_2 ratios, the structural information relates to the orientation of bond vectors relative to the diffusion tensor [20]. To make use of these data, three additional parameters need to be determined: the overall effective rotational correlation time, the diffusion anisotropy and the rhombicity. These parameters can be readily extracted from a histogram of the distribution of measured T_1/T_2 values [32], in an analogous manner to that employed to determine the magnitude of the alignment tensor [27]. In addition, these three parameters can also be optimized using grid search procedures, using the same approach as that employed for dipolar couplings [25].

Long-range distance restraints

The success of short-range (≤5 Å), NOE-derived, interproton distances in determining 3D structures of proteins lies in the simple observation that, in most proteins, there are a significant number of residues far apart in the linear amino acid sequence that are close together in space [1–3]. There are clearly instances, however, in which long-range distance information extending over 15–25 Å could potentially provide very useful, if not critical, data that complement long-range orientational restraints by introducing translational information.

Long-range distance restraints can be obtained from paramagnetic relaxation effects. In the general case, this involves introducing a paramagnetic spin label at an appropriate point(s) in the molecule [33]. In proteins, this can be achieved, for example, by chemically modifying a solvent-accessible cysteine (or, if a suitable cysteine is not present in the molecule of interest, it can be introduced by site-directed mutagenesis) with a nitroxide spin label (such as MTSSL, 1-oxyl-2,2,5,5-tetramethyl-Δ³-pyrroline-3-(methyl)methanethiosulfonate) [15••]) or by adding a copper-binding ATCUN motif (e.g. the sequence Gly–Ser–His) at the N or C terminus of the protein [34••]. For nucleic acids, spin labeling can be achieved by introducing 4-thiouracil or 4-thiothymine at an appropriate location in the sequence and then coupling the sulfur group to a proxyl spin label, such as 3-(2-iodoacetamidoproxyl) [16].

Distance information is obtained by measuring the paramagnetic relaxation enhancement Γ, that is, the difference in either proton longitudinal or transverse relaxation rates measured in the presence and absence of the paramagnetically active spin label [14••,15••,34••]. The value of Γ is directly proportional to r^{-6}, where r is the distance between the unpaired electron on the spin label and the proton of interest. Recent data indicate that distances up to ~30 Å from the spin label can be determined with reasonable accuracy [14••,15••,34••]. Experimentally, paramagnetic relaxation effects can be reliably measured using 2D heteronuclear single quantum coherence (HSQC)-based pulse sequences. These experiments are generally very sensitive and can be rapidly recorded, thereby permitting many distance restraints to be obtained in a single set of experiments.

Long-range distance restraints are also accessible using non-NMR approaches. These include fluorescence resonance energy transfer (FRET) measurements [35,36••] and cross-linking experiments combined with mass spectrometry [37••,38••]. FRET involves a distance-dependent excited state interaction in which the dipole of the acceptor fluorophore resonates with that of the donor fluorophore. Distances in the range of 10–100 Å are accessible by FRET measurements [35,36••]. The disadvantage of FRET, however, is that appropriate chemical modification, coupled with appropriate mutagenesis, needs to be carried out in order to introduce selective labels in a pairwise manner at various locations in the molecule. Chemical cross-linking with linkers of known length, followed by proteolytic digestion and analysis of the fragments by mass spectrometry to establish the site of the cross-links, can also yield useful distance information over a range of ~10–15 Å [37••,38••].

A number of potentially very useful applications of paramagnetic relaxation enhancement (as well as of other methods of long-range distance determination) can be envisaged.

The first application relates to establishing long-range order. A typical case would involve determining the relative orientation of two domains in a situation in which the number of short (<5 Å) interproton distance contacts between the domains is very limited. This is often found to be the case with modular proteins. Even if the few short interproton distance contacts between the two domains can be observed in the NOE spectra, the accumulation of errors over long distances will necessarily result in poor definition of the relative orientation of the two domains. Although long-range orientational information from dipolar couplings can be very useful in this regard, the additional translational information afforded by long-range distance restraints can be very powerful indeed.

The second example that could readily benefit from long-range distance information involves the study of complexes in cases in which it may be difficult to observe intermolecular NOEs. This could be due to the intrinsic lack of sensitivity of heteronuclear separated/filtered NOE experiments, which falls off rapidly with increasing molecular weight, or to sidechain dynamics at the interface, resulting in the broadening of interfacial sidechain resonances and consequent quenching of NOEs. The utility of such an approach has been nicely demonstrated with a protein–RNA complex [16].

The third example, with potentially important ramifications in the field of structural genomics, involves the use of paramagnetic relaxation measurements to speed up the determination of global folds [14••,15••]. The general strategy would involve selectively substituting cysteine residues at various locations in the protein, modifying these cysteines with a spin label and then measuring distances from the spin label to amide or even sidechain protons. Although the feasibility of such an approach has been demonstrated in principle, it may not be quite so trivial to apply in the case of a protein of unknown fold, as the key aspect of the methodology is that the cysteine residues to which the spin label is going to be attached must be solvent accessible (i.e. if one does not wish to perform scanning mutagenesis, one has to already have a reasonable idea of the structure in order to introduce cysteines at sensible locations).

Improvements in simulated annealing

Simulated annealing is a powerful optimization method for finding the global minimum region of a target function characterized by a complex energy surface that includes many local minima. The basic idea is to overcome energy barriers by initially raising the temperature of the system, followed by slow cooling, while exploring configuration space using either Monte Carlo or molecular dynamics sampling. In the case of NMR structure determination, molecular dynamics calculations, based on solving Newton's equations of motion, are generally employed [3].

Molecular dynamics simulations are most commonly carried out in the Cartesian space of atomic positions. In recent years, it has become clear that internal coordinates can be computationally more efficient for molecular structure calculations. As bond lengths and angles are generally taken as fixed, and no information about these features is generally available from the NMR experiments, it follows that, if these coordinates are removed from the calculations, the conformational space to be searched is reduced in size and can be more rapidly sampled. Thus, for a protein of N atoms, there are $3N$ coordinates in Cartesian space, but only approximately $N/3$ torsion angles in internal coordinate space. In addition, the time step required in torsion-angle dynamics to maintain a given level of energy conservation is typically about ten times larger than that required in atomic Cartesian space, because high frequency bond bending and stretching have been removed.

Efficient recursive algorithms for dynamics in internal coordinates [39] have been implemented for torsion-angle dynamics in a number of X-ray and NMR refinement packages [40–43]. More recently, a new implementation of a general internal variable dynamics module (IVM) for efficient molecular dynamics and gradient minimization [44••] has been introduced into the program XPLOR-NIH [43]. This implementation offers a number of unique advantages that open up numerous possibilities for structure refinement. For example, a variety of coordinates can be employed that include not only torsion angles but also more general coordinates, which are appropriate when some degrees of freedom are of interest and others are not. Thus, using the IVM algorithm, it is possible to carry out mixed-mode dynamics: for example, rigid body and Cartesian coordinate dynamics, rigid body and torsion-angle dynamics, and a combination of all three. This is particularly useful in the refinement of a protein–protein complex for which the backbone coordinates of the isolated protein structures are known. Additional features include an efficient integrator with automatic time step adjustment. Finally, the IVM algorithm includes loop constraints to maintain bond lengths in ring topologies while simultaneously permitting full sampling of ring puckering. Examples of the power of the IVM algorithm will be discussed in the following section.

New methods for determining orientations of domains and subunits

The past year has seen significant progress with regard to rapidly solving the structures of protein–protein complexes and efficiently determining relative domain and subunit orientations. All these methods rely on the use of residual dipolar couplings to provide orientational information, supplemented, in the case of protein–protein complexes, by intermolecular NOE data that have both orientational and translational content.

With the advent of structural proteomics, structural studies on protein–protein complexes have become increasingly important. Thus, although the structure of an individual protein can certainly provide insight into the design of further biochemical studies aimed at probing structure/function relationships, it is evident that the structure of a complex yields much greater functional insight. In many instances, protein complex formation involves no significant changes in backbone conformation. Thus, if the structures of the individual proteins are already known at high resolution and it can be shown that the backbone conformation remains essentially unchanged upon complex formation (e.g. by comparison of dipolar coupling data measured for the complex with the X-ray structures of the free proteins), one can then make use of conjoined rigid body/torsion-angle dynamics [44••] to solve the structure of the complex on the basis of intermolecular NOE data and backbone NH dipolar couplings [45••,46••]. In this procedure, only the interfacial sidechains are allowed to alter their conformation. The backbone and noninterfacial sidechains of one protein are held fixed, whereas those of the second protein are only allowed to rotate and translate as a rigid body. This procedure has been applied with considerable success in the case of two complexes: the 40 kDa enzyme I–HPr complex [44••,45••], which had previously been solved by conventional means [47], and the 30 kDa IIA^{Glc}–HPr complex [46••]. It should be emphasized that this approach can readily be extended to cases in which significant changes in backbone conformation are localized to specific regions of the protein, such as the binding interface. In such a case, both the interfacial sidechains and the relevant portions of the protein backbone would be allowed to alter their conformation, and the experimental data would also have to include intramolecular restraints (NOE, dipolar coupling, etc.) relating to that portion of the backbone.

To apply the above approach (or one involving a rigid body systematic grid search) to a multidomain covalently linked system, it is necessary to sever the connection between the two domains [48•–51•]. In doing so, translational information is lost and needs to be reintroduced either in the form of artificial distance restraints [48•] or by only permitting hinge rotations of one domain relative to the other [49•]. This has been applied successfully to a number of cases, including the domain-swapped dimer of the antiviral protein cyanovirin-N [48•], maltose-binding protein [49•], T4 lysozyme [50•] and tRNA [51•].

An alternative and simpler approach can be used that involves the application of a modified form of conjoined rigid body/torsion-angle dynamics, preserving the linker between the two domains [52••]. In this case, each domain constitutes a cluster (i.e. a rigid body) and movement of one domain relative to the other occurs through rotation of backbone torsion angles (ϕ,ψ) within the linker connecting the two domains. As the linkers are typically rather short, the number of degrees of freedom is effectively limited to a very small number of torsion angles (two for each residue in the linker). This approach has been successfully applied to the determination of the solution structure of the domain-swapped dimer of cyanovirin-N [52••].

Rapid protein fold identification

In 1999, it was shown that it is feasible, in principle, to determine the 3D structure of a protein on the basis of backbone dipolar couplings (N-H, N-C′ and HN-C′) combined with a knowledge of secondary structure (which can be readily derived from a qualitative interpretation of NOE data used in conventional assignment procedures) [53]. This has important implications for speeding up NMR structure determination, as the measurement of dipolar couplings only requires knowledge of residue-specific assignments, which can readily be obtained in a sequential manner using through-bond double and triple resonance correlation experiments. In contrast, analysis of NOE spectra is complicated by the fact that ambiguities arising from spectral overlap can make interpretation of

the data problematic, given that correlations are observed between protons close in space and that the most important structural NOEs involve residues that are far apart in the sequence [1,2]. The authors [53] were careful to point out that these results could only be achieved under special circumstances related to topology and secondary structure content. Thus, for the small, 56-residue, GB1 domain of streptococcal protein G, which consists of ~90% secondary structure comprising a four-stranded β sheet on top of which lies a single helix, the accuracy was increased from 4.3 Å using only secondary structure restraints to ≤1 Å using both dipolar coupling and secondary structure restraints [53]. For other proteins with more complex topologies and/or predominantly helical content, it was found that some long-range sidechain–sidechain NOE data were essential to establish the fold correctly [53]. Specifically, they showed that the dipolar couplings had to be supplemented by NH–NH/methyl/aromatic and methyl/aromatic-methyl/aromatic NOEs [53].

Since this early study, some progress has been made using a variety of database approaches, in which fragments of the polypeptide chain are extracted from the crystallographic database on the basis of dipolar couplings and backbone chemical shift data, and are subsequently assembled to produce a complete polypeptide fold [54••,55]. Others have attempted to build and subsequently assemble fragments on the basis of dipolar coupling data alone [56•–58•]. These studies provide proof of principle that it is possible to determine the polypeptide fold of a protein on the basis of a comprehensive set of dipolar couplings alone. In practice, however, such approaches are more likely to be useful as an aid to conventional NMR structure determination, rather than as a method of structure determination in their own right, as they are subject to two rather fundamental limitations. First, the data in most cases are unlikely to be comprehensive and, second, the dipolar couplings do not contain any translational information. This can be crucial for correctly assembling fragments, particularly as the dipolar coupling data permit several alternative solutions.

Another area of development has involved the use of dipolar couplings to recognize protein folds [59,60•,61••]. This may hold significant promise for rapidly establishing a fold that is similar to an existing one in cases in which there is no significant sequence identity between the protein under study and those present in the Protein Data Bank (PDB).

Finally, dipolar couplings, in combination with backbone chemical shift data, torsion-angle database potentials of mean force [23,24] and a potential term for the radius of gyration to ensure appropriate compactness [21], have been used to refine structures on the basis of dipolar couplings alone in cases involving relatively small reorientations of secondary structure elements [62•]. This has been used for examining the structures of the N- and C-terminal

domains of calmodulin in their calcium-ligated [63••] and unligated [62•] states.

Database potentials of mean force to improve non-bonded contacts

Even with extensive experimental data, the accuracy and quality of an NMR structure are dependent on the description of the non-bonded interactions employed. Thus, structures calculated using a purely repulsive van der Waals term have a tendency to be expanded, as, by way of entropic considerations, there are many more expanded than compacted states that satisfy the experimental NMR restraints [21]. As the packing density in proteins is fairly constant, this can be resolved by the inclusion of a radius of gyration restraint [21]. For nonglobular structures, such as nucleic acids, in which the interbase packing density is highly variable and conformation dependent, inclusion of a radius of gyration restraint would simply distort the structure [22••].

One, albeit computationally intensive, approach to improving this situation is to employ a complete empirical energy function including realistic van der Waals and electrostatic interactions, and possibly taking into account water, either explicitly or by means of a generalized Born solvent model [64••,65]. Unfortunately, this is not entirely trivial, particularly for nucleic acids, as the careful balance of the various terms in the empirical energy function is immediately lost upon the inclusion of experimental NMR restraints. In addition, current empirical energy functions are of limited accuracy and only offer an approximate description of physical reality. It is therefore not surprising to find that the use of a full empirical energy function also introduces distortions into NMR structures of nucleic acids, particularly in the form of compression and local structural features intermediate between A and B DNA [22••]. We expect this approach to become more useful, however, as potentials improve and as additional experience is gained.

An alternative approach is to make use of potentials of mean force derived from high-resolution crystal structures that provide a statistical description, in simple geometric terms, of the relative positions of pairs of neighboring bases (both intrastrand and interstrand) in Cartesian space [22••]. The aim of such a potential is similar to that of the widely used torsion-angle database potentials of mean force [23,24], namely to bias sampling during simulated annealing to physically reasonable regions of conformational space within the range of possibilities that are consistent with experimental NMR restraints. The utility of this approach for nucleic acids has recently been demonstrated for a dodecamer [22••] for which extensive NOE and dipolar coupling data had previously been measured [66]. It was shown by cross-validation against independent NMR observables (i.e. both dipolar couplings and NOE-derived interproton distance data) that inclusion of the base–base positioning potential results in a significant increase in

accuracy and obviates artifactual distortions arising from limitations of conventional descriptions of the non-bonded contact term [22••]. Equally importantly, it was shown, using the SRY–DNA complex as an example of a structure which contains a region of DNA that is highly unusual and distorted, that the base–base positioning potential of mean force does not hinder unusual interactions and conformations from being satisfactorily sampled and reproduced [22••,67••]. The same approach can also be used for RNA, as well as to describe sidechain–sidechain interactions in proteins and protein–protein complexes, and sidechain–nucleic acid interactions in protein–nucleic acid complexes.

Conclusions

In this review, we have summarized some recent developments in theoretical and computational aspects of biomolecular NMR spectroscopy aimed at increasing the accuracy, reliability and speed of NMR macromolecular structure determination. These include refinement against novel NMR observables that afford long-range restraints (both orientational restraints, derived from residual dipolar couplings, CSA and heteronuclear relaxation, and translational restraints, obtained by paramagnetic relaxation enhancement measurements, FRET and mass spectroscopy); improvements in simulated annealing methods (in particular, torsion-angle dynamics and the use of conjoined rigid body/torsion-angle dynamics, which promises to be particularly powerful for the rapid elucidation of structures of macromolecular complexes); and the development of new database potentials of mean force aimed at improving non-bonded interactions. Future directions in the field will probably be aimed at increasing the applicability and routine use of these various approaches, which should have a major impact on structural genomics and proteomics.

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. Wüthrich K: *NMR of Proteins and Nucleic Acids.* New York: John Wiley; 1986.

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

3. Clore GM, Gronenborn AM: **New methods of structure refinement for macromolecular structure determination by NMR.** *Proc Natl Acad Sci USA* 1998, **95**:5891-5898.

4. Bax A, Kontaxis G, Tjandra N: **Dipolar couplings in**
•• **macromolecular structure determination.** *Methods Enzymol* 2001, **339**:127-174.
A comprehensive review relating to all aspects of the measurement of dipolar couplings for macromolecules in solution.

5. Prestegard JH, Kishore AI: **Partial alignment of biomolecules: an aid to NMR characterization.** *Curr Opin Chem Biol* 2001, **5**:584-590.

6. Tolman JR: **Dipolar couplings as a probe of molecular dynamics and structure in solution.** *Curr Opin Struct Biol* 2001, **11**:532-539.

7. Tjandra N, Bax A: **Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium.** *Science* 1997, **278**:1111-1114.

8. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A: **Use of dipolar ^{1}H-^{15}N and ^{1}H-^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution.** *Nat Struct Biol* 1997, **4**:732-738.

9. Clore GM, Starich MR, 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, **120**:10571-10572.

10. Hansen MR, Mueller L, Pardi A: **Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions.** *Nat Struct Biol* 1998, **5**:1065-1074.

11. Bax A, Vuister GW, Grzesiek S, Delaglio F, Wang AC, Tschudin R, Zhu G: **Measurement of homo- and heteronuclear couplings from quantitative J correlation.** *Methods Enzymol* 1994, **239**:79-125.

12. Yang D, Kay LE: **Determination of protein backbone dihedral angle ψ from a combination of NMR-derived cross-correlation spin relaxation rates.** *J Am Chem Soc* 1998, **120**:9880-9887.

13. Ottiger M, Bax A: **An empirical correlation between amide deuterium isotope shift effects on ^{13}Cα chemical shifts and protein backbone conformation.** *J Am Chem Soc* 1997, **119**:8070-8075.

14. Battiste JL, Wagner G: **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* 2000, **39**:5355-5365.
This paper presents a detailed study illustrating the potential utility of long-range distance restraints derived from paramagnetic relaxation effects in protein structure determination.

15. Gaponenko V, Howarth JW, Columbus L, Gasmi-Seabrook G, Yuan J,
•• Hubbell WL, Rosevear PR: **Protein global fold determination using site-directed spin and isotope labeling.** *Protein Sci* 2000, **9**:302-309.
This paper illustrates the potential utility of long-range distance restraints derived from paramagnetic relaxation effects in rapidly determining protein folds. This could have a major impact on low-resolution structural studies and accelerate fold identification in structural genomics.

16. Ramos A, Varani G: **A new method to detect long-range protein-RNA contacts: NMR detection of electron-proton relaxation induced by nitroxide spin-labeled RNA.** *J Am Chem Soc* 1998, **120**:10992-10993.

17. Cornilescu G, Bax A: **Measurement of proton, nitrogen and carbonyl chemical shift shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase.** *J Am Chem Soc* 2000, **122**:10143-10154.

18. Wu Z, Tjandra N, Bax A: **^{31}P chemical shift anisotropy as an aid in**
•   **determining nucleic acid structure in liquid crystals.** *J Am Chem Soc* 2001, **123**:3617-3618.
This paper illustrates the impact of direct refinement against ^{31}P CSA restraints in nucleic acid structure determination.

19. Choy WY, Tollinger M, Mueller GA, Kay LE: **Direct structure refinement of high molecular weight proteins against residual dipolar couplings and carbonyl chemical shift changes upon alignment: an application to maltose binding protein.** *J Biomol NMR* 2001, **21**:31-40.

20. 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.

21. Kuszewski J, Gronenborn AM, Clore GM: **Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration.** *J Am Chem Soc* 1999, **121**:2337-2338.

22. Kuszewski J, Schwieters CD, Clore GM: **Improving the accuracy of**
•• **NMR structures of DNA by means of a database potential of mean force describing base-base positional interactions.** *J Am Chem Soc* 2001, **123**:3903-3918.
The authors present a very detailed study in which a novel base–base positional potential of mean force is described. Its impact on the accuracy of nucleic acid structure determination was demonstrated by cross-validation.

23. 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.

24. Kuszweski J, Clore GM: **Source of and solutions to problems in the refinement of protein NMR structures against torsion angle potentials of mean force.** *J Magn Reson* 2000, **146**:249-254.

25. Clore GM, Gronenborn AM, Tjandra N: **Direct refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude.** *J Magn Reson* 1998, **131**:159-162.

26. Tjandra N, Marquardt J, Clore GM: **Direct refinement against**
  •   **proton–proton dipolar couplings in NMR structure determination of macromolecules.** *J Magn Reson* 2000, **142**:393-396.
This paper describes the necessary algorithms required to refine directly against ^{1}H-^{1}H dipolar couplings, which provide not only orientational but also translational restraints.

27. 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.

28. Warren JJ, Moore PDB: **A maximum likelihood method for**
  •   **determining D_A^{PQ} and R for sets of dipolar coupling data.** *J Magn Reson* 2001, **149**:271-275.
This paper describes a reliable maximum likelihood method for deriving the magnitude of the alignment tensor directly from the measured dipolar couplings. This is particularly valuable when the number and distribution of measured dipolar couplings are limited, as in the case of nucleic acids.

29. Moltke S, Grzesiek S: **Structural constraints from residual tensorial couplings in high resolution NMR without an explicit term for the alignment tensor.** *J Biomol NMR* 1999, **15**:77-82.

30. Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S: **An easy way**
  •   **to include weak alignment constraints into NMR structure calculations.** *J Biomol NMR* 2001, **21**:275-280.
This paper illustrates a practical application of the algorithm presented in [29].

31. Shindo H: **NMR relaxation process of ^{31}P in macromolecules.** *Biopolymers* 1980, **19**:509-522.

32. 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* **120**:4889-4890.

33. Kosen PA: **Spin labeling of proteins.** *Methods Enzymol* 1989, **177**:86-121.

34. Donaldson LW, Skrynnikov NR, Choy WY, Muhandiram DR, Sarkar B,
  •• Forman-Kay JD, Kay LE: **Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy.** *J Am Chem Soc* 2001, **123**:9843-9847.
This paper describes a simple approach to introduce a paramagnetic label at the N or C terminus of a protein, and illustrates how structures can be refined directly against paramagnetic relaxation enhancements without any need to convert the measured enhancements into distances.

35. Mendelsohn AR, Brent R: **Protein interaction methods — toward an endgame.** *Science* 1999, **284**:1948-1950.

36. Hillisch A, Lorenz M, Diekman S: **Recent advances in FRET:**
  •• **distance determination in protein–DNA complexes.** *Curr Opin Struct Biol* 2001, **11**:201-207.
A very nice review illustrating the power of FRET for obtaining long-range distance information on protein–DNA complexes.

37. Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID,
  •• Gibson BW, Dollinger G: **High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry.** *Proc Natl Acad Sci USA* 2000, **97**:5802-5806.
This paper presents a groundbreaking approach based on intramolecular cross-linking, proteolytic cleavage and mass spectrometry to derive structural distance restraints, which can be used to determine protein folds. Although it is of limited utility at the present time, this paper provides a proof of principle and, with improved biochemical technology, this approach may have wide-reaching implications for protein structure analysis because only minute quantities of protein are required for mass spectrometry.

38. Krylov DM, Hurley JB: **Identification of proximate regions in a**
  •• **complex of retinal guanyly cyclase 1 and guanylyl cyclase-activating protein-1 by a novel mass spectrometry-based method.** *J Biol Chem* 2001, **276**:30648-30654.
This paper is similar in spirit to [37${}^{••}$], but deals specifically with identifying intermolecular contacts in protein–protein complexes using cross-linking, proteolytic cleavage and mass spectrometry. Once the technology is fully developed, this approach is likely to revolutionize low-resolution structure determination of protein–protein complexes in cases in which the structures of the individual proteins are minimally altered upon complex formation.

39. Jain A, Vaidehi N, Rodrioguez G: **A fast recursive algorithm for molecular dynamics simulation.** *J Comput Phys* 1993, **106**:258-268.

40. Rice LM, Brünger AT: **Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement.** *Proteins* 1994, **19**:277-290.

41. Güntert P, Mementhaler C, Wüthrich K: **Torsion angle dynamics for NMR structure calculation with the new program DYANA.** *J Mol Biol* 1997, **273**:283-298.

42. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunsteleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al.*: **Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination.** *Acta Crystallogr* 1998, **D54**:901-921.

43. Clore GM, Kuszewski J, Schwieters CD, Tjandra N: XPLOR-NIH version 1.1.2 (2002) on the World Wide Web URL: http://nmr.cit.nih.gov/xplor-nih

44. Schwieters CD, Clore GM: **Internal coordinates for molecular**
  •• **dynamics and minimization in structure determination and refinement.** *J Magn Reson* 2001, **152**:288-302.
A detailed presentation of a new IVM incorporated into the XPLOR-NIH program [43]. This module permits completely general hinge definitions to be employed, thereby allowing mixed-mode dynamics to be carried out (i.e. rigid body and Cartesian coordinate dynamics, rigid body and torsion-angle dynamics, or a combination of all three). This is particularly useful when some degrees of freedom are of interest and others are not.

45. Clore GM: **Accurate and rapid docking of protein-protein**
  •• **complexes on the basis of intermolecular nuclear Overhauser enhancement data and dipolar couplings by rigid body minimization.** *Proc Natl Acad Sci USA* 2000, **97**:9021-9025.
This paper demonstrates how the structures of protein–protein complexes can be very rapidly determined on the basis of only a small number of inter-molecular NOEs in combination with dipolar couplings. This paper provided the underlying basis for later and improved methodologies comprising conjoined rigid body/torsion-angle dynamics (see [44${}^{••}$,52${}^{••}$]).

46. Wang G, Louis JM, Sondej M, Seok YJ, Peterkofsky A, Clore GM:
  •• **Solution structure of the phosphoryl transfer complex between the signal transducing protein HPr and IIA^{Glucose} of the** *Escherichia coli* **phosphoenolpyruvate:sugar phospho-transferase system.** *EMBO J* 2000, **19**:5635-5649.
This paper presents the first structure determination of a protein–protein complex using mixed-mode dynamics, in which the protein backbone and noninterfacial sidechains are treated as rigid bodies, and only the interfacial sidechains are allowed to alter their conformation. In addition, this paper nicely illustrates how structurally completely different underlying backbone scaffolds can be used to present very similar interaction surfaces.

47. Garrett DS, Seok YJ, Peterkofsky A, Gronenborn AM, Clore GM: **Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr.** *Nat Struct Biol* 1999, **6**:166-173.

48. Bewley CA, Clore GM: **Determination of the relative orientation of**
  •   **the two halves of the domain-swapped dimer of cyanovirin-N in solution using dipolar couplings and rigid body minimization.** *J Am Chem Soc* 2000, **122**:6009-6016.
This paper illustrates how the relative orientation of two domains can be determined by rigid body minimization using only a small number of backbone N-H dipolar couplings in conjunction with filtering by back calculation of dipolar couplings on the basis of molecular shape. In this instance, the authors also demonstrate unambiguously that the relative orientation of the two halves of the domain-swapped dimer of cyanovirin-N in solution at neutral pH differs by ~80° from that observed in the X-ray structure crystallized at low pH.

49. Skynnikov NR, Goto NK, Yang D, Choi WY, Tolman JR, Mueller GA,
  •   Kay LE: **Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with β-cyclodextrin.** *J Mol Biol* 2000, **295**:1265-1273.
This paper presents a method for determining the relative orientation of protein domains on the basis of dipolar couplings by only permitting hinge rotations of one domain relative to the other.

50. Goto N, Skrynnikov NR, Dahlquist FW, Kay LE: **What is the average**
  •   **conformation of bacteriophage T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR.** *J Mol Biol* 2001, **308**:745-764.
Using the methodology presented in [49${}^{•}$], the authors demonstrate unambiguously that the average orientation of the two domains of T4 lysozyme in solution is slightly different (~10°) from that observed in the crystal structure.

51.   Nollova ET, Hanse MR, Pardi A: **Global structure of RNA**
•   **determined with residual dipolar couplings.** *J Am Chem Soc* 2000,
       122:11561-11562.
This paper illustrates how the orientation of the long and short arms of a
tRNA molecule can be determined from dipolar couplings.

52.   Clore GM, Bewley CA: **Using conjoined rigid body/torsion angle**
••   **simulated annealing to determine the relative orientation of**
       covalently linked protein domains from dipolar couplings. *J Magn*
       Reson 2002, **154**:329-335.
This paper illustrates how the relative orientation of two domains can be
obtained by conjoined rigid body/torsion-angle dynamics on the basis of
dipolar couplings by keeping the linker between the two domains intact and
only permitting backbone torsion-angle degrees of freedom within the linker.
This is illustrated for the case of the domain-swapped dimer of cyanovirin-N,
in which the relative orientation of the two halves of the dimer in solution
at neutral pH differs by ~80° from that observed in the X-ray structure
crystallized at low pH.

53.   Clore GM, Starich MR, Bewley CA, Cai M, Kuszewski J: **Impact of**
       residual dipolar couplings on the accuracy of NMR structures
       determined from a minimal number of NOE restraints. *J Am Chem*
       Soc 1999, **121**:6513-6514.

54.   Delaglio F, Kontaxis G, Bax A: **Protein structure determination using**
••   **molecular fragment replacement and NMR dipolar couplings.**
       J Am Chem Soc 2000, **122**:2142-2143.
This paper presents the first demonstration of a complete determination of a
protein backbone fold using a database fragment search approach, on the
basis of dipolar couplings and backbone chemical shifts.

55.   Zweckstetter M, Bax A: **Single-step determination of protein**
       substructures using dipolar couplings: aid to structural genomics.
       J Am Chem Soc 2001, **123**:9490-9491.

56.   Fowler CA, Tian F, Al-Hashimi HM, Prestegard JH: **Rapid**
•   **determination of protein folds using residual dipolar couplings.**
       J Mol Biol 2000, **304**:447-460.
This paper illustrates how the structures of fragments of a polypeptide chain
can be determined from dipolar couplings and subsequently assembled to
produce a complete polypeptide fold on the basis of minimal NOE data.

57.   Tain F, Valafar H, Prestegard JH: **A dipolar coupling based strategy**
•   **for simultaneous resonance assignment and structure**
       determination of protein backbones. *J Am Chem Soc* 2001,
       123:11791-11796.
This paper shows how a range of heteronuclear backbone dipolar couplings
can be measured using only a ^{15}N-labeled protein and, subsequently, how
these data may be used to generate structures of protein fragments.

58.   Hus JC, Marion D, Blackledge M: **Determination of protein**
•   **backbone structure using only residual dipolar couplings.** *J Am*
       Chem Soc 2001, **123**:1541-1542.
This paper shows that, in principle, a complete backbone fold can be
determined on the basis of a complete set of backbone dipolar couplings
collected in two different alignment media.

59.   Annila A, Aitio H, Thulin E, Drakenberg T: **Recognition of protein**
       folds via dipolar couplings. *J Biomol NMR* 1999, **14**:223-230.

60.   Meiler J, Peti W, Griesinger G: **DipoCoup: a versatile program for**
•   **3D-structure homology comparison based on residual dipolar**
       couplings and pseudocontact shifts. *J Biomol NMR* 2000,
       17:283-294.
This paper presents an algorithm for comparing the structures of homologous
proteins using dipolar couplings.

61.   Andrec M, Du P, Levy RM: **Protein structural motif recognition via**
••   **NMR residual dipolar couplings.** *J Am Chem Soc* 2001,
       123:1222-1229.
A very thorough study showing how protein structural motifs can be recognized
by database searches on the basis of dipolar couplings.

62.   Chou JJ, Li S, Bax A: **Study of conformational rearrangement and**
•   **refinement of structural homology models by the use of**
       heteronuclear dipolar couplings. *J Biomol NMR* 2000, **18**:217-227.
This paper presents a simple method for refining structures using only
dipolar couplings in cases involving small reorientations of secondary
structure elements.

63.   Chou JJ, Li S, Klee CB, Bax A: **Solution structure of Ca^{2+}**
••   **calmodulin reveals flexible hand-like properties of its domains.**
       Nat Struct Biol 2001, **8**:990-997.
The authors, using dipolar couplings, clearly show that the orientations of
the helices in the N-terminal domain of Ca^{2+}-ligated calmodulin differ
significantly from those observed in the crystal structure, providing a clear
example of the potential consequences of crystal packing interactions.

64.   Basford D, Case DA: **Generalized Born models of macromolecular**
••   **solvation effects.** *Annu Rev Phys Chem* 2000, **51**:129-152.
A comprehensive review of the recent progress in empirical energy potentials,
focusing in particular on the generalized Born model as a method for taking
into account water in a computationally highly efficient manner.

65.   Onufriev A, Bashford D, Case DA: **Modification of the generalized**
       Born model suitable for macromolecules. *J Chem Phys* 2000,
       B104:3712-3720.

66.   Tjandra N, Tate S, Ono A, Kainosho M, Bax A: **The NMR structure of**
       a DNA dodecamer in an aqueous dilute liquid crystalline phase.
       J Am Chem Soc 2000, **122**:6190-2000.

67.   Murphy EC, Zhurkin VB, Louis JM, Cornilescu G, Clore GM:
••   **Structural basis for SRY-dependent 46-X,Y sex reversal:**
       modulation of DNA bending by a naturally occuring point
       mutation. *J Mol Biol* 2001, **312**:481-499.
This paper presents a detailed NMR structural study of wild-type SRY and
M9I SRY complexed to DNA. It shows how modern NMR methodology,
including the use of dipolar couplings and various database potentials of
mean force relating to torsion angles and base–base interactions, can be
used to determine the structures of protein–DNA complexes at sufficiently
high accuracy to be able to discern a difference in DNA bend angle of about
15° between the wild-type and mutant complexes. This is the structural
difference that is responsible for 46-X,Y sex reversal.