

# RESIDUAL DIPOLAR COUPLINGS IN NMR STRUCTURE ANALYSIS\*<sup>1</sup>

---

Rebecca S. Lipsitz and Nico Tjandra

Laboratory of Biophysical Chemistry, National Heart, Lung, and Blood Institute,  
National Institutes of Health, Bethesda, Maryland 20892; email: nico@helix.nih.gov;  
lipsitzr@nhlbi.nih.gov

**Key Words** nuclear magnetic resonance, RDCs, structure determination, domain orientation

■ **Abstract** Residual dipolar couplings (RDCs) have recently emerged as a new tool in nuclear magnetic resonance (NMR) with which to study macromolecular structure and function in a solution environment. RDCs are complementary to the more conventional use of NOEs to provide structural information. While NOEs are local-distance restraints, RDCs provide long-range orientational information. RDCs are now widely utilized in structure calculations. Increasingly, they are being used in novel applications to address complex issues in structural biology such as the accurate determination of the global structure of oligonucleotides and the relative orientation of protein domains. This review briefly describes the theory and methods for obtaining RDCs and then describes the range of biological applications where RDCs have been used.

---

\*The U.S. Government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

<sup>1</sup>**Abbreviations:** RDCs, residual dipolar couplings; NMR, nuclear magnetic resonance; DMPC, dimyristoyl-phosphatidyl choline; DHPC, dihexanoyl-phosphatidyl choline; CHAPSO, (cholamidopropyl)dimethylammonio-2-hydroxy-1-propane-sulfonate; IPAP, in-phase, antiphase; CTAB, cetyltrimethylammonium bromide; SDS, sodium dodecyl sulfate; CSA, chemical shift anisotropy; SVD, single value decomposition; CaM, calmodulin; VEGF, vascular endothelial growth factor; Prp40, yeast splicing factor pre-mRNA processing protein 40; BL, barley lectin; TFIIIA, transcription factor IIIA; MalBP, maltodextrin binding protein; CPCl, cetylpyridinium chloride; IRE, iron-responsive element RNA I; AMM,  $\alpha$ -methyl mannoside; ManBPA, mannose binding protein A; LacNAc, acetylactosamine; ACP, acyl carrier protein; MsrA, methionine sulfoxide reductase; *Echm*, *Erwinia chrysanthemi*; SCP, calcium binding protein; CaMKII, CaM kinase II; smMLCK, myosin light chain kinase from smooth muscle; CaMKK, CaM kinase kinase; CaMKI, CaM kinase I; CaMKIV, CaM kinase IV; NOE, nuclear Overhauser effect.

## CONTENTS

INTRODUCTION .....	388
BACKGROUND .....	389
Theoretical Framework .....	389
Measurement of RDCs .....	390
Determination of $A_{\alpha}$ and R .....	391
Data Refinement .....	393
STRUCTURE REFINEMENT AND DOMAIN ORIENTATIONS .....	395
Structural Information from Fitting of RDCs .....	396
Relative Domain Orientations .....	397
DNA/RNA STRUCTURE REFINEMENT .....	401
RDCs from DNA Molecules .....	401
RDCs from RNA Molecules .....	403
CONFORMATION OF SMALL MOLECULES AND BOUND	
LIGANDS .....	404
DETERMINATION OF PROTEIN FOLDS FROM RDCs .....	406
CONCLUSIONS .....	409

## INTRODUCTION

NMR structural biologists are always seeking ways to increase the size limit of biological macromolecules that are amenable to study and to expand the range of biological questions that can be addressed. Recent methods such as TROSY(50) and protein labeling strategies (29) as well as the availability of higher magnetic fields have dramatically increased the size of macromolecules that can be studied by NMR. However, the ability to study larger macromolecules, in of itself, still does not allow one to answer many relevant questions, particularly those pertaining to global structure and domain interactions. This is due to the fact that until recently the principal NMR data for structure determination were the NOE and scalar J couplings, which are entirely local in nature.

RDCs have dramatically altered the types of applications to which NMR methods can be applied. RDCs are complementary to NOEs; they provide orientational information, both short range and long range. Similar to NOEs, RDCs are utilized as restraints in molecular dynamics calculations. In contrast to an NOE, which provides a distance restraint between two atoms, an RDC contains distance information as well as angles formed by a vector connecting the two atoms within a tensor axis system. Reviews have been written that emphasize the theory behind the methods and techniques. However, within the past few years there has been an explosion in the number of systems and problems that have been studied using RDCs; therefore, a review summarizing these applications is warranted. Many of these applications address unresolved structural discrepancies among previous NMR structures, crystal structures, and other biophysical data. Furthermore, several new RDC techniques are useful for facilitating structural genomics efforts (4).

## BACKGROUND

### Theoretical Framework

In NMR, the  $r^{-6}$  dependence of NOEs means that NOEs can usually be detected only between protons within 5 Å. This information is both short range and local; the presence of two pairs of NOEs does not provide any information on how they are related to each other. In contrast, because the dipolar coupling is defined in terms of a molecular coordinate frame, the measurement of two dipolar couplings provides orientational information on how each dipole is related to the molecular coordinate frame and in turn, to each other. This information does not depend on the translational distance between the two dipolar vectors.

The dipolar coupling is a useful phenomenon with which to characterize structure because it depends on distance, orientation, and dynamics. Dipolar couplings have long been a mainstay in solid-state NMR, but recent developments have made them routine in solution NMR. The dipolar coupling is a through-space interaction that arises between any two magnetically active nuclei. As a result of the effects of Brownian motion, dipolar couplings average to zero under isotropic conditions and are only observed under anisotropic conditions. For two dipole-coupled nuclei, A and B, the observable dipolar coupling in solution,  $D_{AB}$ , can be expressed as:

$$D_{AB}(\theta, \phi) = A_a^{AB} \left\{ (3 \cos^2 \theta - 1) + \frac{3}{2} R (\sin^2 \theta \cos 2\phi) \right\}. \quad 1.$$

$A_a^{AB}$  and  $R$  are the axial and rhombic components, respectively, of the molecular alignment tensor,  $\mathbf{A}$ , in the principal coordinate frame.  $\mathbf{A}$  contains the principal components  $A_{xx}$ ,  $A_{yy}$ , and  $A_{zz}$ . According to typical convention, the magnitudes of the principal components are  $|A_{zz}| \geq |A_{yy}| \geq |A_{xx}|$ .  $A_a$  is equal to  $1/3[A_{zz}^{AB} - (A_{xx}^{AB} + A_{yy}^{AB})/2]$  and  $A_r$  is equal to  $1/3[A_{xx}^{AB} - A_{yy}^{AB}]$ .  $A_a$  is in units of hertz and  $R$ , which is equal to  $A_r^{AB}/A_a^{AB}$ , is unitless and always positive.  $\theta$  is the angle between the internuclear bond vector and the  $z$  axis of the alignment tensor,  $\phi$  is the angle between the projection of the internuclear bond vector onto the  $x$ - $y$  plane and the  $x$  axis.  $A_a^{AB}$  is equal to

$$A_a^{AB} = - \left( \frac{\mu_0 h}{16\pi^3} \right) S \gamma_A \gamma_B \langle r_{AB}^{-3} \rangle A_a, \quad 2.$$

where  $\mu_0$  is the permeability in a vacuum,  $h$  is Planck's constant,  $S$  is the generalized order parameter,  $\gamma_A$ ,  $\gamma_B$  are the gyromagnetic ratios of the two nuclei,  $\langle r \rangle$  is the time-averaged internuclear distance, and  $A_a$  is the aforementioned axial component of the molecular alignment tensor. In the applications presented here,  $\langle r \rangle$  between directly bonded nuclei is known and  $S$  is generally assumed to be constant, so the  $\theta$  and  $\phi$  angles are the only variables that contribute to the values of the RDC.

To extract dipolar coupling data, the molecule must behave anisotropically. Otherwise, there is no preferred orientation and the average value of  $D_{AB}$  is 0.

In solution NMR, solutes behave according to Brownian motion and the dipolar interaction averages to zero. To obtain RDCs in solution, a cosolute is needed that causes a partial alignment and a net nonzero average value of  $\theta$  and  $\phi$  without causing severe coupling interactions and distorted spectra. This allows RDCs to be observed while retaining the overall simplicity of solution NMR spectra (65). Dipolar couplings on fully aligned samples such as solids are typically tens of kHz, whereas dipolar couplings from partially aligned solution samples are usually under 100 Hz. RDCs can be observed in molecules that have a sufficiently large magnetic susceptibility anisotropy such as metalloproteins with paramagnetic centers or diamagnetic systems such as DNA where the small anisotropy in each base is additive over the entire molecule. The magnetic susceptibility causes a field-dependent alignment of molecules (30).

## Measurement of RDCs

Liquid crystals for the purposes of alignment in NMR were first introduced in 1963 by Saupe (59, 60) to study small molecules, but the concentrations used led to multiple dipolar couplings for individual nuclei, thus making the spectra difficult to resolve. Bicelles were introduced in the early 1990s and have since been used extensively to achieve a sufficient degree of alignment (55, 56). Bicelles are disk-shaped particles that are made from the detergents DMPC and DHPC, typically in a ratio of 3:1. The concentrations in NMR samples are usually 5% (w/v), but the degree of protein alignment can be “tuned” by adjusting the bicelle concentration. The alignment of bicelles is temperature dependent. At room temperature the bicelles behave isotropically, but at higher temperatures ( $\sim 37^\circ\text{C}$ ) they take on a liquid crystal-type behavior, aligning with their normal perpendicular to the direction of the magnetic field (48). The mechanism by which the neutral bicelles exert their orienting properties is thought to be primarily due to steric hindrance (76). The degree of alignment can be determined by measuring the  $^2\text{H}$  quadrupolar splitting in the HDO resonance. The splitting arises from exchange between isotropic bulk  $\text{H}_2\text{O}$  and aligned  $\text{H}_2\text{O}$  molecules associated with the aligned bicelle. In addition to bicelles, many other types of alignment media and protocol have been developed. These include bicelles using different detergents, phage particles, purple membrane fragments, strain-induced gels, and CPCI/hexanol, among others (21, 32, 35, 57, 58).

In NMR spectra, RDCs appear as an additional contribution to the scalar  $J$  coupling splitting. The magnitude of  $D_{AB}$  can be positive or negative and must be determined by taking the difference of the splitting under anisotropic conditions ( $J + D$ ) and under isotropic conditions ( $J$ ). Methods for measuring RDCs have been described previously (8, 23, 52).

It is possible to alter some of the alignment cosolutes so they produce a different alignment tensor. Bicelles, for example, can be doped with small charged amphiphiles to alter their charge. CTAB confers a positive charge on bicelles, whereas SDS confers a negative charge. In addition, salt and pH can also change

the alignment tensor. The choice of which aligning medium to use is protein dependent and usually determined empirically.

## Determination of $A_a$ and $R$

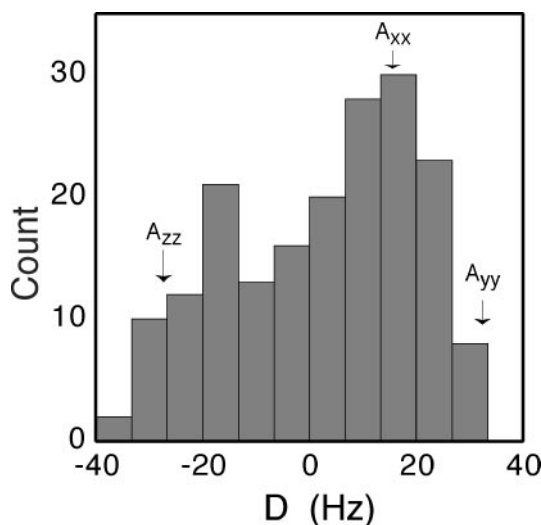
To use RDCs in any type of structure refinement, good estimates for  $A_a$  and  $R$  in Equation 1 must be available. There are several methods for determining  $A_a$  and  $R$ , and the choice of which one to use depends in part on whether a reasonably accurate structure is available prior to the refinement.

In the histogram method demonstrated by Clore et al. (19), the RDCs are measured, normalized to account for the properties of different nuclei, and plotted in a histogram (Figure 1). This histogram closely resembles a CSA powder pattern spectrum characteristic of solid-state NMR spectra, where the values of the chemical shift tensor can be estimated from the pattern. Values for  $A_{zz}$ ,  $A_{yy}$ , and  $A_{xx}$ , are taken from the three extrema of the histogram, shown in Figure 1. These values can be used with Equations 3, 4, and 5 to solve for  $A_a$  and  $R$ .

$$\text{A-B lies along } D_{zz} : \theta = 0^\circ \quad A_{zz} = 2A_a, \quad 3.$$

$$\text{A-B lies along } D_{yy} : \theta = 90^\circ, \phi = 90^\circ \quad A_{yy} = -A_a \left\{ 1 + \frac{3}{2}R \right\}, \quad 4.$$

$$\text{A-B lies along } D_{xx} : \theta = 90^\circ, \phi = 0^\circ \quad A_{xx} = -A_a \left\{ 1 - \frac{3}{2}R \right\}. \quad 5.$$



**Figure 1** A histogram of normalized RDCs,  ${}^1D_{\text{NH}}$ ,  ${}^1D_{\text{C}\alpha\text{H}\alpha}$ ,  ${}^1D_{\text{C}'\text{N}}$ ,  ${}^1D_{\text{C}'\text{C}\alpha}$ , for the KH3 domain from ribonucleoprotein (reprinted with permission from Reference 6).

This method can be used in cases where no previous structural information is available. The key to using this method successfully is that the ensemble of RDCs must sample a wide range of  $\theta$  and  $\phi$ . For many types of biomolecules this is unlikely to be the case.

Another approach put forth by Clore et al. (20) makes use of a grid search to determine  $A_a$  and  $R$ . First,  $A_a$  is estimated by taking the average of the low RDC values:

$$A_a = \frac{-A_{\min}}{1 + 1.5R}. \quad 6.$$

Then a series of short simulated annealing calculations are performed with the value of  $A_a$  from Equation 6 while varying the value of  $R$ . The premise of this method is that a structure refined in a simulated annealing protocol with the correct value of  $A_a$  and  $R$  will have the lowest overall energy.

In cases where a fairly accurate structure is available, SVD can be used to fit RDCs to a series of linear equations to determine the direction cosines that are then used to determine the three principal components of the Saupe order matrix,  $\mathbf{A}_{ij}$  (39). The  $\mathbf{A}_{ij}$  order matrix using Cartesian coordinates is a  $3 \times 3$  matrix in which the subscripts  $i$  and  $j$  refer to the  $x$ ,  $y$ , or  $z$  axes of the alignment tensor. Since the order matrix has the properties that it is both symmetric and traceless, only five parameters are needed to define it:  $A_a$ ,  $R$ , and the three Euler angles  $\alpha$ ,  $\beta$ , and  $\gamma$ , which specify the rotation between the alignment tensor frame and an arbitrary molecular frame. The three principal components of the alignment tensor are determined by diagonalizing the matrix. Since the matrix contains five unknowns, measurement of five individual RDCs within a rigid structure is sufficient to determine the alignment tensor.  $A_a^{AB}$  in Equation 2 contains the generalized order parameter  $S$ , which is assumed to be constant. Thus, if the structural region used for the five measurements is not sufficiently rigid, then the calculated alignment tensor is actually an average alignment tensor. If the input structure is not good enough, it can be difficult to obtain a unique solution for  $A_a$ ,  $R$ , and the Euler angles.

Some of the above methods may be difficult to apply to RNA and DNA. These molecules have a lower proton density and are accompanied by limited spectral dispersion, and because of the high helical content, the RDCs do not always adequately sample enough orientations to obtain unbiased solutions. Warren & Moore (74) have presented a method for determining  $A_a$  and  $R$  specifically for oligonucleotides. This protocol, which is based on the maximum-likelihood method (73), uses the program SSIA to determine  $A_a$  and  $R$  on the basis of the shape of the molecule (76). An initial family of structures is calculated using only NOE and dihedral angle restraints. This family is used as input for SSIA to generate a range of  $R$  values.  $A_a$  is determined for each value of  $R$  by the histogram method. Another family of structures is calculated with RDC restraints using this range of  $R$  and  $A_a$  values. The refined structures are used as input for the SSIA program and this protocol is repeated until the range of values for  $R$  and  $A_a$  converge.

## Data Refinement

For the purpose of using RDCs to calculate NMR structures, RDCs are usually used not in initial structure calculations but rather in a refinement stage of structure calculations. The reasons are that the potential energy surface is very rough and including RDCs initially may trap the structure into a false minimum, leading to convergence problems (45).

A module for incorporating RDCs into structure calculations has been developed for use in XPLOR-NIH (61). This protocol includes a target function in the form of a quadratic harmonic potential,

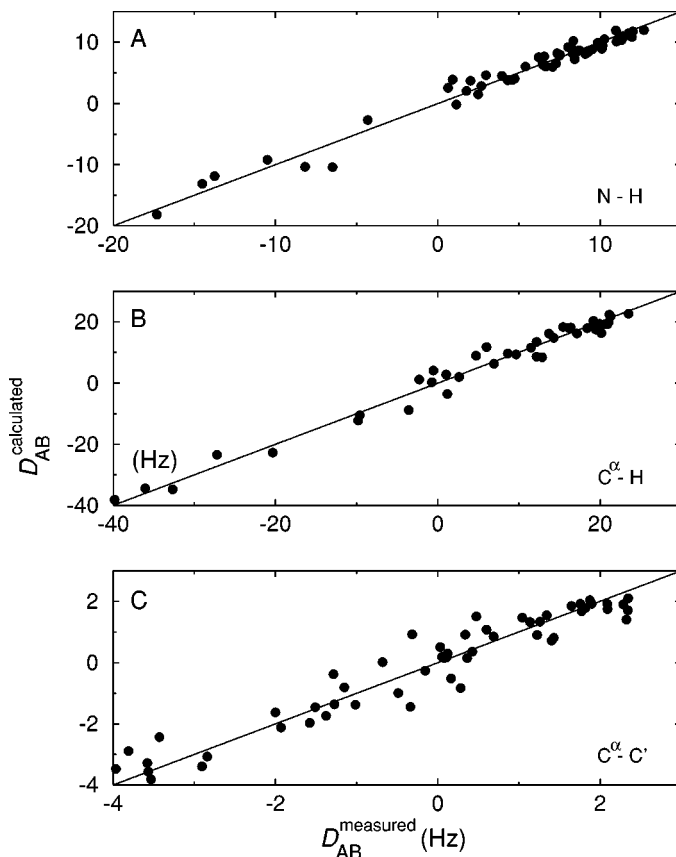
$$E_{dip} = k_{dip}(D^{calc} - D^{meas})^2, \quad 7.$$

where  $D^{calc}$  and  $D^{meas}$  are the calculated and measured RDCs, respectively,  $k_{dip}$  is the force constant, and  $E_{dip}$  is the dipolar energy. In cases where  $S^2$  is less than 0.60, a half-harmonic potential should be used to take into account that the actual value of  $D_{AB}$  is smaller than the measured value of  $D_{AB}$  (18). The force constant should be chosen so that the dipolar RMS is equal to the error in the measured RDCs, approximately 0.2–2.0 Hz.

The alignment tensor is specified by a four-atom pseudo molecule, OXYZ. O represents the center of the molecule and the atoms X, Y, and Z represent the three orthogonal axes of the tensor. The orientation of the alignment tensor with respect to the molecular coordinate frame is determined during the simulated annealing (16). The simulated annealing protocol attempts to shift the bond vector orientations to minimize the difference between measured and calculated RDCs in a manner that is consistent with other experimental data without distorting the covalent structure.

Inclusion of RDC restraints generally improves the precision of families of structures. A common measure of improvement is an increase in the number of residues that falls in the most favored region on a Ramachandran plot. In the case of the DNA binding domain of the transcription factor GATA-1 bound to a 16-bp DNA, the inclusion of RDCs increases the residues found in the most favored regions from 62 to 79% (66). In almost all refinement studies a significant decrease in the backbone RMSD has been noted and quite often RDC restraints lead to decreases in the non-RDC energy terms (6). A plot of the experimental RDCs versus the calculated RDC restraints for the structures prior to and after refinement should show an increase in agreement between the measured and calculated RDCs. An example of this kind of correlation is shown in Figure 2 for ubiquitin. Another option is to perform a cross-validation by omitting one subset of RDCs from the calculations and determine if the agreement between the measured and calculated values for this subset improves.

The Q factor and  $R_{dip}$  are qualitative measures of the agreement between RDCs that are not used in the structure calculation with the other structural restraints used during the simulated annealing calculations. The calculated RDCs can be determined from a refined structure or another structure that is being compared,



**Figure 2** Measured versus calculated RDCs from ubiquitin. (A)  ${}^1D_{\text{NH}}$ , (B)  ${}^1D_{\text{C}\alpha\text{H}\alpha}$ , and (C)  ${}^1D_{\text{C}'\text{C}\alpha}$ . Reprinted with permission from Reference 68.

such as a homologous crystal structure. The Q factor is:

$$Q = \left\{ \frac{\sum_{i=1,N} (D_i^{\text{meas}} - D_i^{\text{calc}})^2}{\sum_{i=1,N} (D_i^{\text{meas}})^2} \right\}^{1/2} \quad 8.$$

$R_{\text{dip}}$  is the same as the Q factor but in a form similar to the crystallographic free R factor (18).

$$R_{\text{dip}} = \left\{ \frac{5 \langle (D_a^{\text{meas}} - D_a^{\text{calc}})^2 \rangle}{[2(D_a^{\text{AB}})^2(4 + 3R^2)]} \right\}^{1/2} \quad 9.$$

Both the Q factor and  $R_{\text{dip}}$  vary from 0 to 1; a lower value indicates better agreement. An NMR structure refined with dipolar couplings could have a Q factor as low as 0.16 (22).

## STRUCTURE REFINEMENT AND DOMAIN ORIENTATIONS

Although the introduction of RDCs into the refinement of macromolecule structures is relatively recent, it is now almost routine to include them in structure calculations. Refinement with RDCs leads to novel insights into NMR structures and in turn into biological function. Often a structure refined with RDCs is significantly different from a previously solved NMR structure without RDCs or the crystal structure, and structures solved with RDCs are helpful for eliminating discrepancies between NMR and X-ray crystallography structures of the same molecule.

The solution conformation of a molecule can also be determined by fitting the RDCs to a structure and then adjusting the structure so that the calculated RDCs closely match the measured RDCs. This is most often done in cases where there is no solution structure of the molecule in question but there is a homologous structure, either a solution structure of a closely related molecule or a crystal structure. Fitting of RDCs is also done routinely when trying to determine the relative domain orientation of multidomain molecules.

The ability to appreciate the unique features of a protein depends on how accurately it is defined. Rat S100B( $\beta\beta$ ) is a member of the  $\text{Ca}^{2+}$ -binding EF hand signaling proteins, for which  $\text{Ca}^{2+}$  binding leads to conformational changes within each EF hand. In the original solution structure of rat apo S100B( $\beta\beta$ ), the position of the third helix was somewhat obscured, thus making it difficult to fully understand the extent of conformational rearrangement upon  $\text{Ca}^{2+}$  binding (25). This was due to a problem that is inherent in typical NOE-based structure determination, which is that there is an error associated with each NOE distance measurement and many small errors in local structure are cumulative and can lead to poorly defined global structure.

After refining the structure of S100B( $\beta\beta$ ) with over 500 RDC restraints, the backbone RMSD improved from 1.04 to 0.29 Å, and the percentage of residues in the most favored region on the Ramachandran plot improved from 76 to 86%. The increased precision in the position of the third helix suggested that the change in the interhelical angle between the third and fourth helix upon  $\text{Ca}^{2+}$  binding is  $90^\circ$  and not  $116^\circ$ , as previously thought.

In the case of a complex of the N terminus of VEGF<sub>11-109</sub> bound to a 19-residue phage-derived peptide antagonist (49), the inclusion of RDCs helped to improve the global precision of the structure. In particular, the protein-peptide intermolecular contacts were better defined. The refinement resulted in an improvement of the backbone RMSD from 0.60 to 0.37 Å. A thorough understanding of the protein-peptide molecular interaction may aid in the development of antagonists smaller

than the peptide used here. The RDC refinement of Prp40 was fundamental in determining whether its two adjacent WW domains were positioned in a manner that would allow simultaneous binding of poly-proline-containing binding partners (75). The refined structure had a decrease in backbone RMSD from 1.14 to 0.55 Å. It showed that each WW domain leans away from the linker region with the two binding sites facing opposite directions, which suggests it would be possible to bind multiple proteins simultaneously. It is clear that structures which have been refined with RDCs are more precise, and that is a prerequisite for developing insights into function.

## Structural Information from Fitting of RDCs

RDCs in the absence of NOEs can be used to determine a structure by fitting RDCs to a homologous structure. This was recently demonstrated on  $\text{Ca}^{2+}$ -ligated CaM, a protein for which several crystal structures were available, but no solution structure (15). A linker that becomes flexible upon addition of  $\text{Ca}^{2+}$  connects the N and C domains of CaM. The plasticity of the linker is key in allowing recognition and binding to over 100 different target proteins. The only restraints used were a complete set of RDC restraints ( $^1\text{D}_{\text{NH}}$ ,  $^1\text{D}_{\text{C}\alpha\text{H}\alpha}$ ,  $^1\text{D}_{\text{C}'\text{C}\alpha}$ ,  $^2\text{D}_{\text{C}'\text{H}\alpha}$ , and  $^1\text{D}_{\text{C}'\text{N}}$ ) and heteronuclear J couplings to obtain the sidechain conformations.

The limited data used in the simulated annealing were sufficient to reveal substantial differences between the solution structure and a 1 Å crystal structure of  $\text{Ca}^{2+}$ -ligated CaM (PDB entry 1EXR). In the C-terminal domain the relative position of helix V differs by 15°. In the N-terminal domains there are significant deviations in the relative positions of helix I (26°) and helix IV (22°), which result in a solution conformation that is closer to the apo form of CaM in which the hydrophobic binding cleft is less exposed. These changes are significant; yet they would not have been detected had the structure been calculated with NOEs, since back-calculations show that no NOE differs by more than 0.5 Å between the crystal structure and the RDC-refined solution structure of  $\text{Ca}^{2+}$ -ligated CaM.

Crystal structures of the same protein can result in structural discrepancies, making it difficult to determine the physiologically relevant structure. To determine if a particular crystal structure is similar to the solution conformation, measured RDC data can be fit to the crystal structure. In the case of hemoglobin, crystal structures established many years ago the existence of two allosterically regulated conformations: a tense state (T), which has low oxygen affinity, and a relaxed state (R), which has high oxygen affinity. More recently a second conformation of the R state, R2, was crystallized. RDCs were used to determine if the solution conformation of the R state corresponds to either R crystal structure (40).

Initial fitting of the measured RDCs to the hemoglobin tetramer demonstrated that the solution conformation was in a conformation different from either the R or R2 crystal structures and that the conformational differences stemmed from the arrangement of the  $\alpha_1\beta_1/\alpha_2\beta_2$  dimers. These conformational differences are a

result of slight changes in intermolecular contacts and would unlikely be detected with NOE data.

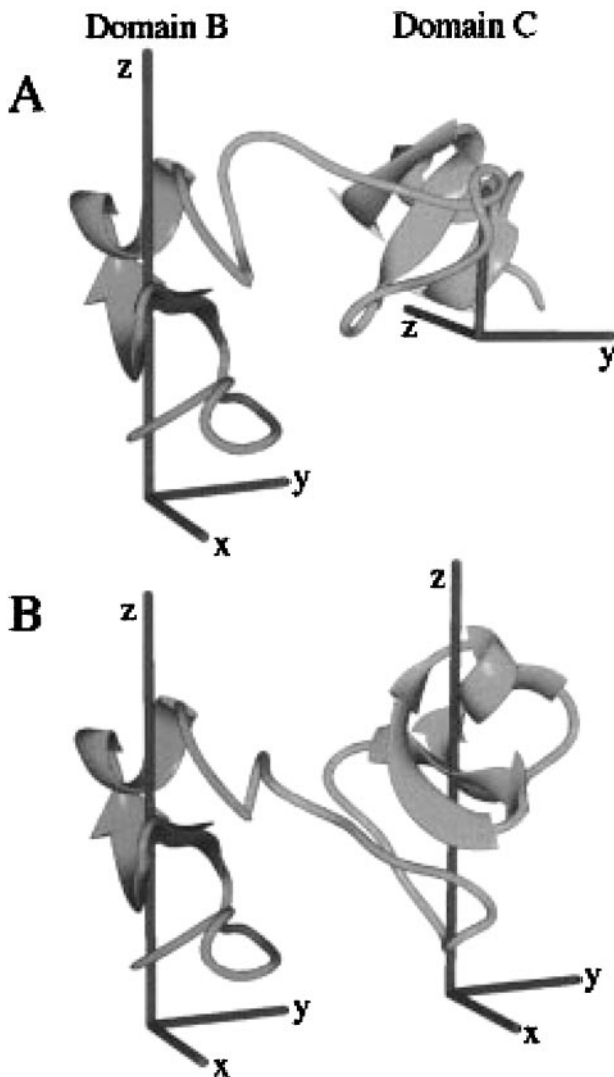
The exact relative orientation of the  $\alpha_1\beta_1/\alpha_2\beta_2$  dimers was determined by taking advantage of the molecule's C2 symmetry. By definition, one of the molecule's principal alignment tensor axes must coincide with the C2 axis (11). Because the  $A_{xx}$  principal axis of the alignment tensor resides midway between the C2 symmetry axes of the R and R2 crystal structures, the relative orientation of the  $\alpha_1\beta_1/\alpha_2\beta_2$  dimers in solution must lie between that of the two crystal structures. The orientation in solution was determined by rotating one dimer relative to the other, recalculating the RDCs after each rotation. The correct orientation of the solution conformation is one that results in the lowest RMSD between the measured and calculated RDCs. If hemoglobin did not have C2 symmetry, a rigid-body simulated annealing protocol in which each domain is held rigid by synthetic restraints would have been necessary.

## Relative Domain Orientations

Most domain orientation procedures are based on the premise that both domains align to the same extent and therefore have a common alignment tensor. The alignment tensor is then determined independently for each domain and the domains are then rotated until the alignment tensors coincide. These protocols do not provide information on the relative translational distance, but rigid-body simulated annealing using other experimental NMR restraints can be used to determine the distance. One of the first demonstrations of domain orientation was performed on the B and C domains from BL, using SVD to calculate the alignment tensor for each domain (27). A previous structure determination of the B and C domains was not able to determine their relative orientation owing to a lack of interdomain NOEs. One possible solution was obtained by rotating the z and x axes of the alignment tensor of domain C by 90° about its y axis. The starting structure of BL and the new orientation are shown in Figure 3.

Anytime RDCs are used, there are four possible orientations that are consistent with the measured values; the  $\cos^2 \theta$  dependence in Equation 1 means that an inversion of  $\theta$  with respect to any of the axes of the principal alignment tensor will result in the same RDC value. Measuring RDCs in a second alignment media that is sufficiently different so that the macromolecule has a different alignment tensor can reduce this orientational degeneracy (53). This is not always necessary because in some cases only one of the possible orientations is consistent with other structural information.

Using a somewhat different method, Tsui et al. (70) determined the relative domain orientation for the three zinc fingers in TFIID when bound to DNA. Previous crystal and solution NMR structures of TFIID showed similarities in the local structure; however, the difference in the relative orientation of the first and second zinc fingers was considerable, 32°. The entire molecule was refined using Equation 1 with the assumption that only the correct relative domain orientation



**Figure 3** The axes system corresponds to the molecular alignment tensor. (A) The domain orientation of a homologous protein, wheat germ agglutinin in the crystal structure. (B) The domain orientation of BL after rotating the alignment tensor of the C domain (reprinted with permission from Reference 30).

would minimize the difference between the measured and calculated RDCs for all three zinc fingers. In the refined structure the relative orientation of the first and second zinc fingers was more precisely defined and closer to the crystal structure than what was seen in the previous NMR structure. However, the difference in the relative orientation between the crystal and solution states was still substantial,  $\sim 20^\circ$ , suggesting that the zinc fingers can adopt different conformations.

Because the same protein can have different relative domain orientations depending on the ligand, information about domain orientation provides insight into function. This has been demonstrated from several studies using MalBP in the apo state, bound to  $\beta$ -cyclodextrin, which is an inactive ligand, and MalBP bound to a natural ligand, maltotriose (26, 63). MalBP is a 370-amino-acid protein that is important in chemotaxis. It comprises two domains roughly equal in size that are connected by a hinge region. The substrate binding site is at the bottom of a cavity formed by the interface of the two domains, and both domains participate in binding peptidoglycan ligands. Crystal structures of MalBP with various ligands differ in the degree to which the two domains are open or closed, and these orientations may not be the same in their solution structure counterparts.

RDCs were used to determine the hinge rotation about the domain linker regions that results in the correct relative domain orientation. In the case of MalBP in the apo form and bound to maltotriose, rigid-body simulated annealing was also performed to determine the interdomain translational information. The accuracy of the hinge rotation method was confirmed by starting with one crystal structure of MalBP in an open conformation and using calculated RDCs from a closed crystal structure to correctly determine the hinge rotation necessary to convert the conformation from open to closed. The solution conformation of MalBP bound to  $\beta$ -cyclodextrin was found to be  $11^\circ$  more closed than the corresponding crystal structure. The solution structure conformation suggests that the N-terminal domain of MalBP may interact more directly with  $\beta$ -cyclodextrin than was indicated from the crystal structure. In contrast to the studies of MalBP bound to  $\beta$ -cyclodextrin, the relative domain orientation of apo MalBP and MalBP bound to maltotriose agreed well with the corresponding crystal structure. The RDC-refined structures for different states of MalBP are shown in Figure 4).

It is sometimes the case that the relative domain orientation in solution represents an average of two conformations. One common scenario is when two distinct conformations, a and b, rapidly interconvert. If one assumes that the open and closed crystal structures of MalBP represent the a and b species and that the relative populations of a and b are  $p_a$  and  $(1 - p_a)$ , respectively, then the measured RDCs are related to the RDCs of the two forms as follows:

$$D_{AB} = p_a D_{AB}^a + (1 - p_a) D_{AB}^b. \quad 10.$$

When the measured RDCs and alignment tensors for the open and closed conformations are used, the above equation reduces to the following:

$$\frac{p_a D_a^a}{(1 - p_a) D_a^b} = 1.67. \quad 11.$$

If one assumes that  $D_a^a \sim D_a^b$ , then  $p_a \sim 0.63$ . In fact, this value is similar to the values determined from the ratio of closure angles of the apo form of MalBP and the form of MalBP bound to maltotriose.

Domain orientation studies have been performed on T4 lysozyme to determine its solution state conformation and to investigate factors that may influence the results of the domain orientation procedure (31). Crystal structures of T4 lysozyme without any substrate show that it is in a closed conformation; yet it is believed that the domain interface must be wider to accommodate peptidoglycan substrates. The refined structures show that the relative domain orientation of the N and C domains in solution are more open than suggested by the crystal structure.

To determine if imprecision in the intradomain atomic coordinates, i.e., structural noise, influences the determination of the hinge rotation used for reorientation, the T4 lysozyme RDCs were fit to nine different crystal starting structures. It was determined that intradomain structural noise in T4 lysozyme results in a variation of hinge rotation axes of  $\pm 2^\circ - \pm 11^\circ$ . T4 lysozyme was also used to test whether the choice of an alignment medium influences the outcome of the refinement. RDCs were measured in four different media. The pair-wise RMSDs for the four reoriented structures ranged from 0.7 to 2.6 Å, which confirms that the medium itself does not exert any influence on the conformation of the protein.

As the previous studies have demonstrated, a prerequisite for domain orientation with RDCs is knowledge regarding any interdomain motion. Braddock et al. (14) have shown how one can detect interdomain motion using RDCs without detailed knowledge of the structure. This method is based on the premise that if a protein consists of multiple domains with little interdomain motion, the system can be described by a single alignment tensor, but if there are substantial interdomain rigid-body motions, each domain will experience a different degree of alignment. For each domain the normalized RDCs are plotted in a histogram. If the profile for the domains is similar, it indicates that the domains share a common alignment tensor, whereas if the profiles are different, it indicates that the domains have different alignment tensors resulting from interdomain motion.

Recently, Clore & Bewley (17) showed how a conjoined rigid-body/torsion angles simulated annealing protocol can be used to determine domain orientations in cyanovirin N, a domain-swapped dimer. This method is based on one used to determine the structure of protein-protein complexes and, similar to the previous methods using hinge rotations, is meant to preserve the interdomain translational information. Each domain is treated as a rigid body, and changes in relative domain position occur by modifying the  $\phi$  and  $\psi$  torsion angles in the linker region. They tested whether the choice in linker length influences the final results and found that the  $\phi$  and  $\psi$  torsion angles in the linker region were similar regardless of the linker length. The  $\phi$  and  $\psi$  angles with the largest changes were concentrated in two residues.

These examples demonstrate the potential information that can be obtained from determining the relative domain orientation of proteins, either through fitting of RDCs or through structure refinement with RDCs. Furthermore, these methods do not give rise to structural artifacts that can arise when using other biophysical techniques.

## DNA/RNA STRUCTURE REFINEMENT

Structure determination of DNA and RNA molecules using the traditional NOE/J scalar coupling approach is inherently more difficult than it is for proteins. Oligonucleotides, especially DNA, are long and cylindrical and often lack the elaborate tertiary structure found in proteins. In addition, they are less proton dense than proteins. Therefore, the number of obtainable NOEs from oligonucleotides is comparatively small. Because most proton-proton distances less than 5 Å are typically only found between adjacent base pairs, there are usually few long-range NOEs. As a result, the local structure tends to be defined relatively well but the global structure suffers. To compound matters, structures of oligonucleotides solved by X-ray crystallography are highly subject to crystal packing forces and may lead to misinterpretations of the global features. Therefore, the orientational information contained in RDCs can have an enormous impact on the accuracy and precision of oligonucleotide structures.

### RDCs from DNA Molecules

One of the first DNA RDC structure refinements was the palindromic Dickerson dodecamer (67). This study established the utility of employing RDCs in DNA structure refinement and highlighted many of the concerns that are unique to nucleic acids. The structure was refined primarily with RDC restraints. A smaller number of NOE restraints were also used, mainly to generate initial structures to which the RDCs could be applied. Artificial backbone torsion angle restraints were necessary to get the final structures to converge. One difference between oligonucleotides and proteins is that the distribution of bond vector orientation is much more restricted in the former. Therefore, a combination of the histogram and grid search approaches was used to determine  $A_a$  and  $R$ . First, a histogram from the  $^1D_{CH}$  values was used to estimate the lower limit of  $A_a$  and then a grid search was used to determine the exact values of  $A_a$  and  $R$ .

Base pair opening, minor groove width, and DNA bending were sensitive to the values of  $A_a$  and  $R$  used. Structures that were calculated with slight variations in  $A_a$  and  $R$  had substantial RMSDs between them. In the crystal structure there were small kinks present in the terminal base pairs that were not present in the refined structure. These kinks give rise to a bend in the crystal structure that is less prominent in the solution structure.

A-tract DNA molecules have been the focus of several studies to determine the impact of RDCs on DNA global structure. A-tract DNA consist of 4 to 6

adenines whose presence results in pronounced curvature of the DNA. The A-tract's irregular structure is important in regulation of transcription. Crystal packing forces make it uncertain whether the curvature seen in crystal structures of A-tract DNA molecules is accurate. Furthermore, conflicting structural and biochemical data have made it difficult to develop a consensus regarding what model best describes the curvature; the improvement in global precision in parameters such as base inclination and tilt has helped to address these issues. RDCs can be used to uniquely define irregular DNA structure, since curvature over several base pairs results in nonuniform CH bond vector orientations. NOE distances, on the other hand, are not significantly affected by DNA curvature.

Pardi and coworkers (72) demonstrated that RDC constraints improve both local and global structure of A-tract DNA molecules. Simulated RDC restraints were used to refine a 10-bp DNA duplex whose structure was previously solved by X-ray crystallography and an extended 14-mer version. Refinement led to significant decreases in the average RMSD to the target structure for both the 10-mer (1.7 to 0.8 Å) and 14-mer (2.7 to 1.2 Å). RMSDs between individual base pairs in the 10-mer and 14-mer showed that RDCs improved the local structure in both molecules equally well. Helical parameters that are difficult to define with NOEs such as helical rise and helical twist were significantly improved. Incorporation of RDCs caused the overall curvature in the 10-mer structure to decrease from 34° to 6°, which was similar to the curvature in the target structure.

This study provided several insights concerning oligonucleotide structure refinement with RDCs. First, to get good convergence and reduce the number of RDC violations, it is necessary to start with a small force constant and to use many small time steps during the molecular dynamics simulations. This is due to the fact that the improvement in global structure is a result of cumulative changes in local structure. As many other groups have subsequently recognized, it is advantageous to start with a structure that is well-defined locally because then the only way to accommodate the RDC data will be to make adjustments to the global structure. The A-tract bending is induced by slowly reorienting the CH vector angles as opposed to making changes in base stacking or other types of local interactions.

Two other recent studies have helped to clarify the discrepancies among A-tract structures and between individual structures and the results of biochemical data regarding A-tract curvature. Macdonald et al. (41) demonstrated that adding RDC restraint to other NMR restraints for a decamer leads to an increase in the calculated A-tract curvature by 10°. The direction of curvature in the RDC-refined structure of the A<sub>6</sub> dodecamer is in agreement with results from biochemical data and contrasts with results from crystal structures. Barbic et al. (7) refined an A<sub>4</sub> decamer with <sup>1</sup>D<sub>CH</sub> RDCs. Although they were more accurately able to define base inclination and roll, both of which contribute to the overall A-tract curvature, they found that the structure was slightly overwound. This may be due to the fact that the CH bond vectors lie mostly parallel to the principal axis of the alignment tensor, and as a result the redundancy in <sup>1</sup>D<sub>CH</sub> restraints defines the helical twist somewhat incorrectly.

In higher-order DNA structures the presence of both inter- and intrastrand NOEs makes it difficult to determine the relative position of individual motifs. Al-Hashimi et al. (3) used RDC restraints to determine the relative orientation of a G·(C-A) triad and a G-tetrad within each monomer of a C<sub>2</sub> symmetric homodimer. Within the homodimer these motifs exist as a quadruplex. Interest in the structure of DNA quadruplexes has increased, as they have increasingly been found among sequences associated with disease (37). Two of the four possible orientations are correct owing to the inherent C<sub>2</sub> symmetry. The relative orientation of the triad and the tetrad was determined with RDCs measured in Pf1 phage and also from field-dependent differences due to the magnetic susceptibility of the DNA at 500 and 800 MHz. Four possible relative orientations were determined for the triad and the tetrad by superimposing the alignment tensors calculated from the phage data. Then the alignment tensor was calculated for the two motifs as though they were one entity by using the field-dependent data. Only the two orientations that are consistent in the phage data and the field-dependent data are correct. This approach should be applicable not only to oligonucleotides, but also to other dimeric systems such as protein dimers.

## RDCs from RNA Molecules

RDCs have also been applied to RNA structure refinement such as the polyadenylation regulatory RNA element bound to the human U1A protein (9). Refinement of this structure led to a change in the relative position of the RNA helices. The global structure of RNA molecules improves upon inclusion of RDCs because RNA often contains intricate and/or irregular structural features. This was demonstrated by determining the relative position of the acceptor and anticodon arms in tRNA<sup>Val</sup> from *Escherichia coli* (47). The correct relative orientation of the two arms was determined by searching for the orientation that minimizes the difference between the calculated and experimental RDCs. In three previous crystal structures of tRNA<sup>Val</sup> there was no agreement in the interarm angle, but the orientation determined from RDCs closely matches that of one of the crystal structures.

The refinement of the hammerhead ribozyme is a dramatic example of differences in RNA stem orientation in the crystal and solution structures. The hammerhead ribozyme contains three stem regions surrounding a catalytic core. Biochemical studies suggest that addition of Mg<sup>2+</sup> initiates folding from an extended conformation into the compact, catalytically active conformation. However, crystal structures both with and without Mg<sup>2+</sup> of the hammerhead solved were in the compact conformation. For comparison, the conformation of hammerhead ribozyme in the absence of Mg<sup>2+</sup> was determined using RDCs (13).

The relative orientations were determined by fixing the position of stem II while rotating stems I and III as rigid bodies to minimize the difference between the measured and calculated RDCs. Because two relative orientation angles were determined, this produced 16 (4 × 4) possible orientations. The coordinates for

each of these structures were used as input for molecular dynamics simulations in XPLOR to incorporate the interstem translational distance information. Thirteen of the structures could be eliminated because they did not agree with NOE data. Of the three remaining structures, all satisfied the NOE data but only one did not have any backbone steric clashes. The conformation of this molecule, unlike the corresponding crystal structure, is elongated with stems I and II extending from opposite sides of the core region. A comparison of these structures is shown in Figure 5.

The IRE is an example of an RNA with irregular structure that was defined much more precisely after including RDCs (44). The IRE contains two stem regions separated by a bulge at the C7 position. The bulge distorts the regular RNA structure and the relative orientation of the helical stems and the hairpin loop is not well defined by NOEs and J couplings. An accurate description is important because the position of the bulge within the IRE often determines its ability to be recognized by proteins involved in iron regulation.

Different refinement strategies were tested on the IRE, suggesting that, in general, unless the local structure is very good to begin with, a better result will occur if the local structure is refined before the global structure. Many local features in the IRE were improved after the refinement including the precision of the glycosidic angles and the sugar puckers. However, several loop nucleotides were still ill-defined after the refinement, suggesting that a range of conformations is energetically consistent with the measured RDCs. The bulge itself appeared unstructured in previous studies, but after RDC restraints were used, most structures in the family clearly showed a unique conformation.

## CONFORMATION OF SMALL MOLECULES AND BOUND LIGANDS

One area of ever-increasing interest in structural biology is ligand-protein interactions and drug design. NMR methods such as SAR by NMR (62) are used to screen for ligands that interact with a given protein. RDCs can complement these methods by determining the average orientation of the bound ligand with respect to the protein. This can be especially beneficial when it is difficult to obtain chemical shift assignments for the nuclei in the ligands that interact with the protein or in cases where ligand hydroxyl groups are in fast exchange with water. Establishing the orientation of a ligand with respect to a protein may make it possible to design and tune new ligands with modified binding properties. RDCs can be used to quickly compare these modified ligands. In addition, RDCs have been used to improve the structure of small ligands that have rigid substructures, such as oligosaccharides, in which it is difficult to establish the conformation using NOEs (38, 43).

The first demonstration using RDCs to determine the conformation of a bound ligand was demonstrated on AMM bound to ManBPA, a 53-kDa protein with threefold symmetry that is involved in innate immune response (12). ManBPA is a

homotrimer with a threefold axis of symmetry that intersects a central coiled-coil region. Each monomer binds one molecule of AMM.

AMM binds weakly to ManBP and is in fast exchange between the bound and unbound forms; the measured ligand RDCs on the ligand-protein complex reflect an average degree of alignment from the free and bound orientations. This is similar to the two-state model for multidomain proteins between the open and closed forms. In this case, the two states are free and bound forms of AMM:

$$D_{AB} = p_{free} D_{AB}^{free} + p_{bound} D_{AB}^{bound}. \quad 12.$$

$p_{free}$  and  $p_{bound}$  can be determined from the binding constants and the molar ratios of AMM and ManBPA in the sample. The RDCs must also be measured in the alignment media for AMM alone. This leaves just one unknown,  $D_{AB}^{bound}$ , in Equation 12. Once the RDCs for AMM in the protein-bound conformation have been determined, the alignment tensor can be calculated and rotated to coincide with the protein's alignment tensor.

Bolon et al. (12) determined the relative orientation of AMM and ManBPA without measuring any RDCs for the protein. Because ManBPA has threefold symmetry, it follows that the alignment tensor axis of highest order,  $A_{zz}$ , must coincide with the protein's threefold symmetry axis (1). Therefore, after the alignment tensor of bound AMM was calculated, its relative orientation to ManBPA was determined by making the  $A_{zz}$  axis of its alignment tensor coincide with the C3 symmetry axis of ManBPA. Because the orientation of AMM relative to ManBPA differed from what was found in the crystal structure of an analogous complex, the alignment tensor for ManBPA was explicitly calculated and the orientation of a similar oligosaccharide, trimannoside, relative to ManBPA was determined by superimposing their alignment tensors (36). The orientation of trimannoside was similar to the orientation for AMM.

These types of studies provide insight into the conformational changes that take place in the protein upon ligand binding. Umemoto et al. (71) examined the lectin protein Galectin-3 in the absence and presence of the disaccharide LacNAc. LacNAc binds in a groove formed by three  $\beta$ -strands. RDC measurement for the bound and free forms of Galectin-3 indicates that the NH bond vector orientations differ between the crystal structure and the solution structure by  $2^\circ$  to  $10^\circ$ . An examination of the RMSD values for RDCs measured on the solution conformation and for RDCs calculated from a crystal structure suggests there may be small conformational differences between the unliganded and liganded solution structures of Galectin-3. Both the unliganded and LacNAc-bound forms of Galectin-3 were refined with RDCs. In the absence of ligand, the binding site cleft contracts somewhat. The two loop residues that are in van der Waals contact with the LacNAc, Asn-166 and Trp-181, move several angstroms upon ligand binding.

Tian et al. (64) have used RDCs to investigate the conformation of trimannoside and the differential motion at the glycosidic linkages. Each of the three pyranose rings that make up trimannoside was considered a separate rigid molecular fragment. In the structure of trimannoside, ring III is situated between ring I and

ring II. Using the RDCs to derive the generalized degree of order, which in turn is used to assess the extent of motion experienced by a rigid fragment, it was determined that internal motion between rings III and I is restricted but that inter-ring motion between rings III and rings II is more significant. As a result, the relative orientation of rings III and I can be determined by superimposing their alignment tensors. Doing the same for rings III and II is not possible, as it would result in a motionally averaged orientation because the alignment tensors are not equivalent.

## DETERMINATION OF PROTEIN FOLDS FROM RDCs

Identifying NOEs is an extremely time-consuming endeavor. Although it is difficult to calculate a protein structure with only RDCs as the experimental restraints, they can be used to expedite this process by determining a protein's fold. Determination of the protein fold is in of itself quite valuable, especially in the era of proteomics, in which determination of the fold of an unknown protein often yields the first clues regarding its function. Some of the concepts behind protein fold determination are useful for other problems such as identifying different ligand binding conformations.

Hus et al. (33) have developed a method to determine the protein fold without using NOEs. Instead they use RDCs and three types of orientational restraints whose information content is analogous but complementary to RDCs: paramagnetic pseudocontact shifts, Curie-dipolar cross-correlated relaxation effects, and heteronuclear relaxation rates. This entire set of restraints can be obtained only on proteins with a paramagnetic center. All the restraints are combined and used for simulated annealing. This method was tested on cytochrome *c*. During the first stage, the preliminary fold is determined by using the restraints only for the secondary structure elements. A subset of structures that best agrees with the experimental data is used for the second stage of simulated annealing, which includes restraints for loop regions. The various tensors are allowed to float during the first phase but are held fixed during the second stage. Although the number of calculated structures whose energy is below the chosen threshold value was low, these structures are accurate. The RMSD between the ensemble of solution structures and a crystal structure of cytochrome *c* is 2.05 Å.

Another approach to determining protein folds with RDCs has been presented by Fowler et al. (28) on ACP. The underlying principle is that within a single domain protein all helices align to the same extent and have a common alignment tensor. Therefore, the helices will have the correct relative orientation if they are rotated so their individual alignment tensors coincide. Backbone assignments were made on each protein and the regions of secondary structure elements were identified with NOEs. Each helix of ACP was designated as an individual fragment. In cases where more RDCs can be measured, a smaller region can be used as a fragment. The smaller the fragment used, the more likely it may be to observe helix curvature or other deviations in local structure. A minimal number of backbone to side chain NOEs were included, since this method does not provide any translational

information. Of the four possible orientations for each fragment, only one structural arrangement was consistent with the NOE data. The final RMSD for the ensemble of structures of ACP was 2.5 Å, and a comparison to a previously solved NMR solution structure of ACP shows that the relative orientation determined with RDCs was fairly accurate.

A conceptually similar method was used to determine the structure of MsrA from *Erwinia chrysanthemi* (MsrA<sup>Echmi</sup>) with backbone and limited side chain assignments and without NOEs (10). First, the tertiary fold of MsrA<sup>Echmi</sup> was evaluated by comparing measured RDCs with calculated RDCs from the crystal structure of MsrA<sup>Ecoli</sup>. For the one region of the protein for which significant differences were found, P196-L202, the structure was constructed by the program *meccano*, which builds a peptide chain from a succession of oriented peptide fragments (34). Each individual peptide plane was considered a rigid structural entity. The alignment tensor was determined for each peptide plane and then the peptide planes were rotated to make their alignment tensors coincide. The peptide was then docked onto the crystal structure of MsrA<sup>Ecoli</sup>. In the new structure the P196-L202 peptide adopted a conformation similar to that in the same region in MsrA<sup>Ecoli</sup>. However, the position of the catalytically important C200 C $\alpha$ -C $\beta$  bond vector points into the protein core, positioning it closer to its disulfide bond partner, whereas in the crystal structure, it points away from the protein core.

Many structure prediction programs have improved their success rates by incorporating RDCs. Rohl & Baker (54) have incorporated RDCs in the ab initio structure prediction program RosettaNMR. In Rosetta, structure prediction is based on the idea that the identities and order of amino acids often dictate the likely secondary structure. Rosetta takes short sequences of the query protein and examines the secondary structure of analogous sequences in the PDB to construct a fragment library. The fragments that are retained in the final library are determined by using a scoring function that takes into account agreement with a multiple sequence alignment, experimental chemical shifts, and RDCs. Monte Carlo-simulated annealing is used to construct a protein model from the small sequence libraries. The contact order, which is the number of amino acids separating contacting residues, improves substantially when both chemical shifts and RDCs are used. Without experimental data, the results were limited to proteins in which the contact order is less than 20, but with the RDCs and chemical shift data, a structure with a contact order of 32 was accurately predicted.

Fold-recognition programs using RDCs have been introduced to determine the topology of a protein with unknown structure (5). This approach was demonstrated on the protein calerythrin, a 20-kDa EF-hand protein for which the structure had not been determined at that time. A search was performed for similar RDCs among six EF-hand proteins only using <sup>1</sup>D<sub>NH</sub> RDCs. These proteins represented four different protein folds and their sequence similarity to calerythrin was less than 30%. As the lengths of secondary structure elements can vary, the comparison starts by using sequence homology to determine which regions of the query and target proteins

to use for the comparison. Low RMSDs were found only between calerythrin and sarcoplasmic calcium binding protein (SCP) from two different organisms. These results are rather striking given that the sequence identity between calerythrin and the two SCPs is 27 and 15%.

Homologous RDCs can be used to identify the  $\phi$  and  $\psi$  backbone dihedral angles that are associated with a particular conformation that can then be used to construct a model of the protein (24). The protein is broken into overlapping seven-residue fragments and the RDCs from each fragment are compared to calculated RDCs from seven-residue fragments in high-resolution crystal structures in the PDB. Twenty matches are selected for each fragment on the basis of a  $\chi^2$  value that is calculated from the agreement between the measured and calculated RDCs and, to a lesser extent, from chemical shift agreement. This process is repeated for every possible seven-residue fragment in the protein, yielding 100 pairs of  $\phi$ ,  $\psi$  angles for each residue. A protein model is then calculated on the basis of these derived  $\phi$ ,  $\psi$  angles. Cases where the spread in  $\phi$ ,  $\psi$  angles from the PDB matches is not narrow indicate that alternative conformations might satisfy the same RDCs, and those  $\phi$ ,  $\psi$  angles should be used with caution. A model of ubiquitin determined this way had a backbone RMSD of 1.2 Å from the crystal structure.

In cases of proteins that have multiple binding partners, the binding mode can be established with  $^1\text{D}_{\text{NH}}$  RDCs in lieu of binding assays or full-structure determinations. RDCs were used to differentiate among three possible binding modes between CaM and two CaM kinases (42). Despite overall similarities in the mechanism of activation of CaM kinases, the position of two important hydrophobic residues on the CaM kinases is critical in classifying the mode of CaM binding. To date, there are three recognized binding modes based on the position of these hydrophobic residues. Sequence alone predicts the mode of binding reasonably well, but RDCs provide a more robust analysis in cases where there are more than two potential hydrophobic anchoring residues.

RDCs are sufficiently sensitive to allow one to discern different modes of binding by the same protein, as seen for CaM. The binding mode of CaM with three different CaM kinases, CaM kinase II (CamKII), myosin light chain kinase from smooth muscle (smMLCK), and CaM kinase kinase (CaMKK), was established from crystal structures. Each of these kinases is recognized by CaM in a different manner. The binding modes of CaM to two kinases whose structures have not been determined, CaM kinase I (CaMKI) and CaM kinase IV (CaMKIV), was determined by measuring  $^1\text{D}_{\text{NH}}$  RDCs for CaM in complex with peptides from the kinases. The measured CaM RDCs were plotted against calculated RDCs from each of the three crystal structure complexes mentioned above. In both cases, the correlation coefficient was only greater than 0.95 when correlated with the RDCs from CaM bound to smMLCK. This indicates that the binding mode of CaM with both CaMKI and CaMKIV is the same as for smMLCK. The  $R^2$  values for correlations with either CaMKII or CaMKK ranged from 0.87 to 0.92 Å. The significant difference in  $R^2$  values demonstrates that RDCs can easily and quickly distinguish the mode of peptide binding.

## CONCLUSIONS

A novel structure of a macromolecule often produces many questions, particularly when the information it provides differs from results from other studies. When new techniques were introduced six years ago to obtain dipolar couplings on macromolecules in solution, the potential for refining protein structures was immediately obvious. However, it has since been demonstrated that the range of biological questions RDCs can successfully address is exceptionally broad and ever increasing. RDCs provide the structural biologist with a tool that can be easily implemented to answer some of these questions. Furthermore, as the emphasis of structural genomics shifts from determining a macromolecule's fold to understanding how the nonstatic elements of a structure such as changes in conformation and the ability to recognize other macromolecules promote cellular communication, RDCs will become indispensable. Although this article has focused on the structural applications, researchers are also beginning to exploit RDCs in solution NMR for their dynamics information content. These studies have established a framework to determine interfragment motion, to calculate amplitudes of interdomain motion, and to separate the dynamic contribution to the measured RDC to determine the effective values of  $\theta$  and  $\phi$  (46, 51, 68, 69). Because the dynamics information from RDCs extends to the millisecond timescale, these studies will complement dynamics information from  $^{15}\text{N}$  spin-relaxation experiments.

## ACKNOWLEDGMENTS

We would like to thank Cheryl Hawkins and Reiko Ishimia for careful reading of the manuscript. We would also like to thank James Baber and Professors Lewis E. Kay, Arthur Pardi, and James H. Prestegard for providing figures.

**The Annual Review of Biophysics and Biomolecular Structure is online at <http://biophys.annualreviews.org>**

## LITERATURE CITED

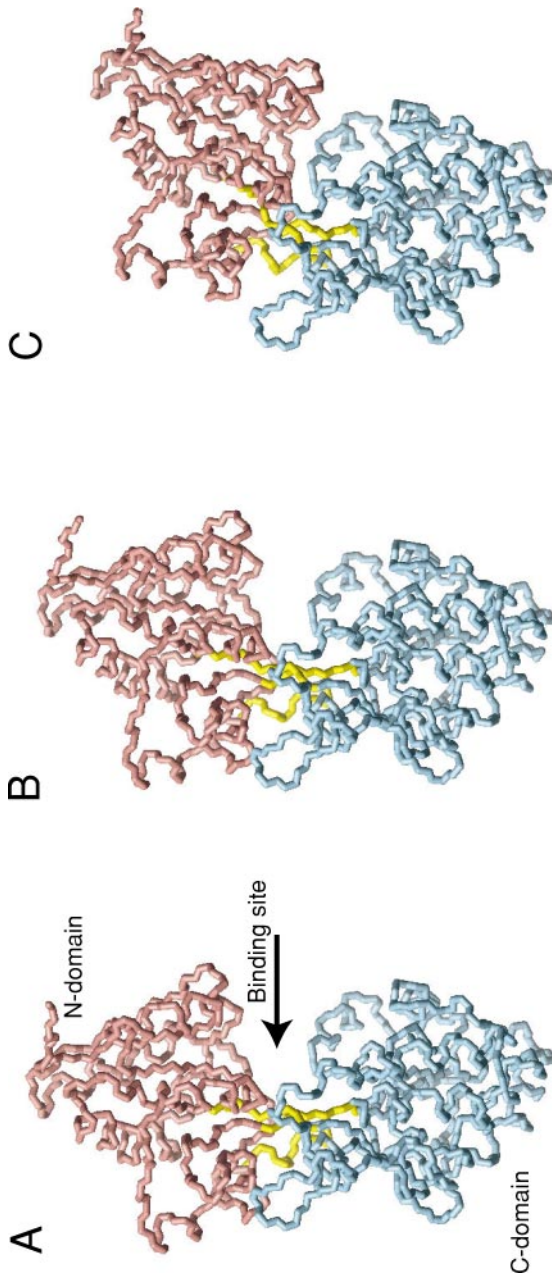
1. Al-Hashimi HM, Bolon PJ, Prestegard JH. 2000. Molecular symmetry as an aid to geometry determination in ligand protein complexes. *J. Magn. Reson.* 142:153–58
2. Al-Hashimi HM, Gosser Y, Gorin A, Hu W, Majumdar A, Patel DJ. 2002. Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J. Mol. Biol.* 315:95–102
3. Al-Hashimi HM, Majumdar A, Gorin A, Kettani A, Skripkin E, Patel DJ. 2001. Field- and phage-induced dipolar couplings in a homodimeric quadruplex: relative orientation of G(C-A) triad and G-tetrad motifs and direct determination of C2 symmetry axis orientation. *J. Am. Chem. Soc.* 123:633–40
4. Al-Hashimi HM, Patel DJ. 2002. Residual dipolar couplings: synergy between NMR and structural genomics. *J. Biomol. NMR* 22:1–8

5. Annala A, Aitio H, Thulin E, Drakenberg T. 1999. Recognition of protein fold via dipolar couplings. *J. Biomol. NMR* 14:223–30
6. Banci L, Bertini I, Savellini GG, Romagnoli A, Turano P, et al. 1997. Pseudocontact shifts as constraints for energy minimization and molecular dynamics calculations on solution structures of paramagnetic metalloproteins. *Proteins Struct. Funct. Genet.* 29:68–76
7. Barbic A, Zimmer DP, Crothers DM. 2003. Structural origins of adenine-tract bending. *Proc. Natl. Acad. Sci. USA* 100:2369–73
8. Bax A, Kontaxis G, Tjandra N. 2001. Dipolar couplings in macromolecular structure determination. *Methods Enzymol.* 339:127–74
9. Bayer P, Varani L, Varani G. 1999. Refinement of the structure of protein-RNA complexes by residual dipolar coupling analysis. *J. Biomol. NMR* 14:149–55
10. Béraud S, Bersch B, Brutscher B, Gans P, Barras F, Blackledge M. 2002. Direct structure determination using residual dipolar couplings: reaction-site conformation of methionine sulfoxide reductase in solution. *J. Am. Chem. Soc.* 124:13709–15
11. Bewley CA, Clore GM. 2000. 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.* 122:6009–16
12. Bolon PJ, Al-Hashimi HM, Prestegard JH. 1999. Residual dipolar coupling derived orientational constraints on ligand geometry in a 53 kDa protein-ligand complex. *J. Mol. Biol.* 293:107–15
13. Bondensgaard K, Mollova ET, Pardi A. 2002. The global conformation of the hammerhead ribozyme determined using residual dipolar couplings. *Biochemistry* 41:11532–42
14. Braddock DT, Cai M, Baber JL, Huang Y, Clore GM. 2001. Rapid identification of medium- to large-scale interdomain motion in modular proteins using dipolar couplings. *J. Am. Chem. Soc.* 123:8634–35
15. Chou JJ, Li S, Klee CB, Bax A. 2001. Solution structure of Ca<sup>2+</sup>-calmodulin reveals flexible hand-like properties of its domains. *Nat. Struct. Biol.* 8:990–97
16. Choy WY, Tollinger M, Mueller GA, Kay LE. 2001. 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* 21:31–40
17. Clore GM, Bewley CA. 2002. Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings. *J. Magn. Reson.* 154:329–35
18. Clore GM, Garrett DS. 1999. R-factor, free R, and complete cross-validation for dipolar coupling refinement of NMR structures. *J. Am. Chem. Soc.* 121:9008–12
19. Clore GM, Gronenborn AM, Bax A. 1998. 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.* 133:216–21
20. Clore GM, Gronenborn AM, Tjandra N. 1998. Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J. Magn. Reson.* 131:159–62
21. Clore GM, Starich MR, Gronenborn AM. 1998. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J. Am. Chem. Soc.* 120:10571–72
22. Cornilescu G, Marquardt JL, Ottiger M, Bax A. 1998. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* 120:6836–37

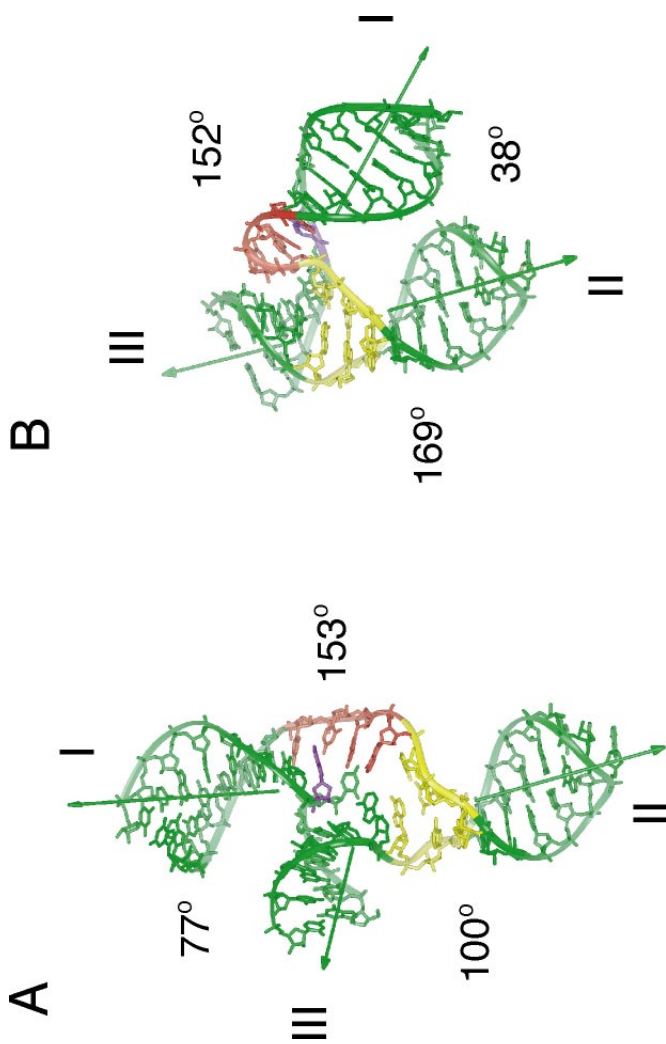
23. de Alba E, Tjandra N. 2002. NMR dipolar couplings for the structure determination of biopolymers in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* 40:175–97
24. Delaglio F, Kontaxis G, Bax A. 2000. Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J. Am. Chem. Soc.* 122:2142–43
25. Drohat AC, Tjandra N, Baldissari DM, Weber DJ. 1999. The use of dipolar couplings for determining the solution structure of rat apo-S100B ( $\beta\beta$ ). *Protein Sci.* 8:800–9
26. Evenäs J, Tugarinov V, Skrynnikov NR, Goto NK, Muhandiram R, Kay LE. 2001. Ligand-induced structural changes to maltodextrin-binding protein as studied by solution NMR spectroscopy. *J. Mol. Biol.* 309:961–74
27. Fischer MWF, Losonczi JA, Weaver JL, Prestegard JH. 1999. Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry* 38:9013–22
28. Fowler CA, Tian F, Al-Hashimi HM, Prestegard JH. 2000. Rapid determination of protein folds using residual dipolar couplings. *J. Mol. Biol.* 304:447–60
29. Gardner KH, Kay LE. 1998. The use of  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 27:357–406
30. Gayathri C, Bothnerby AA, Vanzijl PCM, Maclean C. 1982. Dipolar magnetic-field effects in NMR-spectra of liquids. *Chem. Phys. Lett.* 87:192–96
31. Goto NK, Skrynnikov NR, Dahlquist FW, Kay LE. 2001. What is the average conformation of bacteriophage of T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR. *J. Mol. Biol.* 308:745–64
32. Hansen MR, Mueller L, Pardi A. 1998. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* 5:1065–74
33. Hus JC, Marion D, Blackledge M. 2000. *De novo* determination of protein structure by NMR using orientational and long-range order restraints. *J. Mol. Biol.* 298:927–26
34. Hus JC, Marion D, Blackledge M. 2001. Determination of protein backbone structure using only residual dipolar couplings. *J. Am. Chem. Soc.* 123:1541–42
35. Ishii Y, Markus MA, Tycko R. 2001. Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel. *J. Biomol. NMR* 21:141–51
36. Jain NU, Noble S, Prestegard JH. 2003. Structural characterization of a mannose-binding protein-trimannoside complex using residual dipolar couplings. *J. Mol. Biol.* 328:451–62
37. Keniry MA. 2001. Quadruplex structures in nucleic acids. *Biopolymers* 56:123–46
38. Kiddle GR, Homans SW. 1998. Residual dipolar couplings as new conformational restraints in isotopically  $^{13}\text{C}$ -enriched oligosaccharides. *FEBS Lett.* 436:128–30
39. Losonczi JA, Andrec M, Fischer MWF, Prestegard JH. 1999. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* 138:334–42
40. Lukin JA, Kontaxis G, Simplaceanu V, Yuan Y, Bax A, Ho C. 2003. Quaternary structure of hemoglobin in solution. *Proc. Natl. Acad. Sci. USA* 100:517–20
41. Macdonald D, Herbert K, Zhang X, Polgruto T, Lu P. 2001. Solution structure of an A-tract DNA bend. *J. Mol. Biol.* 306:1081–98
42. Mal TK, Skrynnikov NR, Yap KL, Kay LE, Ikura M. 2002. Detecting protein kinase recognition modes of calmodulin by residual dipolar couplings in solution NMR. *Biochemistry* 41:12899–906
43. Martin-Pastor M, Bush CA. 2001. Refined structure of a flexible heptasaccharide using  $^1\text{H}$ - $^{13}\text{C}$  and  $^1\text{H}$ - $^1\text{H}$  NMR residual dipolar couplings in concert with NOE

- and long range scalar couplings constants. *J. Biomol. NMR* 19:125–39
44. McCallum SA, Pardi A. 2003. Refined solution structure of the iron-responsive element RNA using residual dipolar couplings. *J. Mol. Biol.* 326:1037–50
  45. Meiler J, Blomberg N, Nilges M, Griesinger C. 2000. A new approach for applying residual dipolar couplings as restraints in structure elucidation. *J. Biomol. NMR* 16:245–52
  46. Meiler J, Prompers JJ, Peti W, Griesinger C, Brüschweiler. 2001. Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. *J. Am. Chem. Soc.* 123:6098–107
  47. Molloy ET, Hansen MR, Pardi A. 2000. Global structure of RNA determined with residual dipolar couplings. *J. Am. Chem. Soc.* 122:11561–62
  48. Ottiger M, Bax A. 1998. Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. *J. Biomol. NMR* 12:361–72
  49. Pan B, Li B, Russell SJ, Tom JY, Cochran AG, Fairbrother WJ. 2002. Solution structure of a phage-derived peptide antagonist in complex with vascular endothelial growth factor. *J. Mol. Biol.* 316:769–87
  50. Pervushin K, Riek R, Wider G, Wuthrich K. 1997. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* 94:12366–71
  51. Peti W, Meiler J, Brüschweiler R, Griesinger C. 2002. Model-free analysis of protein backbone motion from residual dipolar couplings. *J. Am. Chem. Soc.* 124:5822–33
  52. Prestegard JH, Al-Hashimi HM, Tolman JR. 2000. NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q. Rev. Biophys.* 33:371–424
  53. Ramirez BE, Bax A. 1998. Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. *J. Am. Chem. Soc.* 120:9106–7
  54. Rohl CA, Baker D. 2002. De novo determination of protein backbone structure from residual dipolar couplings using Rosetta. *J. Am. Chem. Soc.* 124:2723–29
  55. Sanders CR, Hare BJ, Howard KP, Prestegard JH. 1994. Magnetically-oriented phospholipid micelles as a tool for the study of membrane-associated molecules. *Prog. Nucl. Magn. Reson. Spectrosc.* 26:421–44
  56. Sanders CR, Schwonek JP. 1992. Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR. *Biochemistry* 31:8898–905
  57. Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S. 2000. Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR* 18:303–9
  58. Sass J, Cordier F, Hoffman A, Cousin A, Omichinski JG, et al. 1999. Purple membrane induced alignment of biological macromolecules in the magnetic field. *J. Am. Chem. Soc.* 121:2047–55
  59. Saupe A, Englert G. 1963. High-resolution nuclear magnetic resonance spectra of oriented molecules. *Phys. Rev. Lett.* 11:462–64
  60. Saupe A. 1968. Recent results in the field of liquid crystals. *Angew. Chem. Int. Ed.* 7:97–111
  61. Schwieters CD, Kuszewski JJ, Tjandra N, Core GM. 2003. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* 160:65–73
  62. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. 1996. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274:1531–34
  63. Skrynnikov NR, Goto NK, Yang D, Choy

- WY, Tolman JR, et al. 2000. Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with  $\beta$ -cyclodextrin. *J. Mol. Biol.* 295:1265–73
64. Tian F, Al-Hashimi HM, Craighead JL, Prestegard JH. 2001. Conformational analysis of a flexible oligosaccharide using residual dipolar couplings. *J. Am. Chem. Soc.* 123:485–92
65. Tjandra N, Bax A. 1997. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278:1111–14
66. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. 1997. Use of dipolar  $^1\text{H}$ - $^{15}\text{N}$  and  $^1\text{H}$ - $^{13}\text{C}$  couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Biol.* 4:732–38
67. Tjandra N, Tate S, Ono A, Kainosho M, Bax A. 2000. The NMR structure of a DNA dodecamer in an aqueous dilute liquid crystalline phase. *J. Am. Chem. Soc.* 122:6190–200
68. Tolman JR. 2002. A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. *J. Am. Chem. Soc.* 124:12020–30
69. Tolman JR, Al-Hashimi HM, Kay LE, Prestegard JH. 2001. Structural and dynamic analysis of residual dipolar coupling data for proteins. *J. Am. Chem. Soc.* 123:1416–24
70. Tsui V, Zhu L, Huang TH, Wright PE, Case DA. 2000. Assessment of zinc finger orientations by residual dipolar coupling constants. *J. Biomol. NMR* 16:9–21
71. Umemoto K, Leffler H, Venot A, Valafar H, Prestegard JH. 2003. Conformational differences in liganded and unliganded states of Galectin-3. *Biochemistry* 42:3688–95
72. Vermeulen A, Zhou H, Pardi A. 2000. Determining DNA global structure and DNA bending by application of NMR residual dipolar couplings. *J. Am. Chem. Soc.* 122:9638–47
73. Warren JJ, Moore PB. 2001. A maximum likelihood method for determining  $D_a^{PQ}$  and R for sets of dipolar coupling data. *J. Magn. Reson.* 149:271–75
74. Warren JJ, Moore PB. 2001. Application of dipolar coupling data to the refinement of the solution structure of the Sarcin-Ricin loop RNA. *J. Biomol. NMR* 20:311–32
75. Wiesner S, Stier G, Sattler M, Macias MJ. 2002. Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor 40. *J. Mol. Biol.* 324:807–22
76. Zweckstetter M, Bax A. 2000. Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc.* 122:3791–92



**Figure 4** The relative domain orientation of MalBP in three different states: (A) apo-state, (B) bound to  $\beta$ -cyclodextrin, and (C) bound to maltotriose. Reprinted with permission from Elsevier in Reference 29.



**Figure 5** The conformations of the hammerhead ribozyme are shown with the values of the interstem angles. (A) Solution conformation using RDC data in the absence of  $Mg^{2+}$ . (B) The crystal structure in the presence of  $Mg^{2+}$ . Reprinted with permission from Elsevier in Reference 13.