

# The importance of being ordered: improving NMR structures using residual dipolar couplings

Angela M. Gronenborn\*

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Received 11 June 2002; accepted 27 August 2002

Presented by Stuart J. Edelstein

---

**Abstract** – Residual dipolar couplings arise from small degrees of alignment of molecules in a magnetic field. Most biomolecules lack sufficient intrinsic magnetic susceptibility anisotropies for practical purposes; however, alignment can be achieved using dilute aqueous phospholipid mixtures, colloidal suspensions of rod-shaped viruses, complex phases of surfactant systems and strained gels. The stability of the liquid crystalline phases varies with respect to temperature range, pH variation and time and is critically dependent on sample composition and experimental conditions. The magnitude of the residual dipolar couplings depends upon the degree of ordering and allows the determination of the corresponding inter-nuclear vectors with respect to the molecule's alignment frame. Inclusion of dipolar constraints into NMR structure calculations leads to improved precision and accuracy of the resulting structures, especially in cases where the information content provided by traditional NOE constraints is limited. In addition, rapid evaluation of backbone protein folds and determination of the relative orientations of individual components in multi-molecular complexes have become feasible. Dipolar coupling based strategies may well emerge as the most critical developments, in establishing NMR as a valuable and competitive methodology in the structural genomics initiative. *To cite this article: A.M. Gronenborn, C. R. Biologies 325 (2002) 957–966.* © 2002 Académie des sciences / Éditions scientifiques et médicales Elsevier SAS

**NMR / alignment / anisotropy / liquid crystal / residual dipolar couplings / orientational constraints / protein structure determination**

**Résumé** – De l'importance d'être ordonné : l'amélioration des structures RMN à l'aide des couplages dipolaires résiduels. Des couplages dipolaires résiduels se produisent à partir de faibles degrés d'alignement des molécules dans un champ magnétique. La plupart des biomolécules manquent d'anisotropies de susceptibilité magnétique intrinsèque suffisantes pour des applications pratiques ; néanmoins, l'alignement peut être réalisé en utilisant des mélanges aqueux dilués de phospholipides, de suspensions colloïdales de virus en bâtonnets, de phases complexes de systèmes surfactants et de gels contraints. La stabilité des phases cristallines liquides varie avec le domaine de température, le pH et le temps, et dépend de manière critique de la composition de l'échantillon et des conditions expérimentales. L'amplitude des couplages dipolaires résiduels dépend du degré d'ordre et permet la détermination des vecteurs inter-nucléaires correspondants, selon le système d'alignement de la molécule. L'inclusion de contraintes dipolaires dans les calculs de structure RMN conduit à une précision et à une exactitude améliorées des structures résultantes, particulièrement dans le cas où le contenu informationnel procuré par les contraintes NOE traditionnelles serait limité. De plus, une rapide évaluation des replis des motifs protéiques et la détermination des orientations relatives des composants individuels dans les complexes multi-moléculaires sont devenues réalisables. Des stratégies basées sur le couplage dipolaire pourraient bien émerger comme les développements les plus critiques,

---

\*Correspondence and reprints.

E-mail address: gronenborn@nih.gov (A.M. Gronenborn).

en établissant la RMN comme une méthodologie compétitive en génomique structurale. *Pour citer cet article* : A.M. Gronenborn, C. R. Biologies 325 (2002) 957–966. © 2002 Académie des sciences / Éditions scientifiques et médicales Elsevier SAS

**RMN / alignement / anisotropie / cristal liquide / couplages dipolaires résiduels / contraintes d'orientation / détermination de la structure des protéines**

---

## 1. Introduction

Traditionnellement, la détermination de trois-dimensionnelles structures de molécules biologiques emploie des contraintes de distance et d'angle de torsion extraites de NOE et de données de couplage  $J$ , respectivement [1, 2]. Une limitation inhérente à la méthode couramment utilisée de RMN provient de la nature strictement locale des contraintes expérimentales. Heureusement, malgré cet inconvénient, la détermination de structure par RMN pour les protéines globulaires a été extrêmement réussie, principalement parce que les nombreuses distances inter-proton courtes entre acides aminés éloignés dans la séquence sont fortement corrélées, rendant leur conformation extrêmement restrictive.

Néanmoins, l'utilisation de seulement des informations de distance courtes peut limiter l'exactitude des structures dérivées de RMN, en particulier pour les architectures non-globulaires où l'erreur cumulée peut devenir significative ou dans des cas où seuls quelques contacts sont disponibles entre les éléments structuraux. Des exemples de tels systèmes incluent les protéines modulaires et multi-domaines et les acides nucléiques linéaires. Récemment, la méthodologie de RMN a été développée pour exploiter l'alignement faible des molécules dans le champ magnétique, soit causé par l'anisotropie de susceptibilité magnétique propre de la molécule ou en utilisant des milieux cristallins liquides dilués [3]. Cela permet de mesurer les couplages dipolaires résiduels, qui contiennent des informations sur l'orientation d'un vecteur inter-noyau par rapport au tenseur d'anisotropie de susceptibilité magnétique. Ces couplages peuvent être utilisés pour extraire des contraintes angulaires pour être utilisées dans les calculs de structure basés sur la RMN. L'incorporation de ces contraintes supplémentaires dans les algorithmes de détermination de structure conventionnels conduit à d'importantes améliorations des structures de RMN, à la fois localement et globalement.

## 2. Background

L'existence de couplages dipolaires résiduels induits par le champ magnétique pour les protéines en solution isotropique était connue depuis un certain nombre d'années, mais leur utilité pour les études structurales, en particulier, n'a été réalisée qu'après l'apparition de champs magnétiques élevés et de méthodes hétéronucléaires qui ont permis leur détermination précise de façon largement disponible.

Un champ magnétique plus élevé était important, car la taille des couplages dipolaires à une liaison résiduelle varie avec le carré du champ magnétique et de nouvelles expériences ont dû être conçues pour extraire ces couplages extrêmement faibles avec une précision d'une fraction d'herz [4, 5]. Étant donné le faible degré d'alignement des protéines diamagnétiques dans le champ magnétique, résultant en des couplages dipolaires résiduels de typiquement  $< 0.2$  Hz, la faisabilité d'extraire de tels couplages de façon fiable semblait limitée. Plus prometteur, d'un autre côté, est apparu l'alignement induit par le champ magnétique pour les protéines paramagnétiques, les acides nucléiques et les complexes protéine/acide nucléique, qui présentent des anisotropies de susceptibilité magnétique supérieures à  $-10 \times 10^{-34}$  m<sup>3</sup> par molécule, produisant des couplages dipolaires résiduels de  $\sim 0.5$  Hz pour les vecteurs N–H et de  $\sim 0.9$  Hz pour les vecteurs C–H.

Le principal percée en ce qui concerne l'utilisation potentielle des couplages dipolaires pour la détermination de structure biomoléculaire a été la démonstration que des degrés élevés d'alignement moléculaire pouvaient être obtenus en plaçant la molécule sous investigation dans une phase diluée, aqueuse, cristalline liquide de phosphatidylcholine (DHPC) et de dimyristoyl phosphatidylcholine (DMPC) [3, voir 6, 7 pour revues]. Des degrés élevés d'alignement ont été obtenus de cette manière, résultant en des couplages dipolaires à une liaison de 5–40 Hz qui sont facilement détectables en mesurant simplement la séparation dans les spectres HSQC couplés 2D et 3D. Cela a ouvert la voie à de nouvelles expériences pour extraire d'autres types de couplages dipolaires résiduels et a conduit à une série d'activités explorant différents milieux d'alignement avec l'objectif d'améliorer le système bicelle initial ainsi que de découvrir de nouveaux milieux. Plusieurs de ces milieux sont décrits ci-dessous.

Malgré le fait que certains milieux cristallins liquides présentent une viscosité macroscopique élevée par rapport à l'eau pure, de façon intéressante, la diffusion rotationnelle de la molécule biologique dissoute n'est affectée que marginalement. Cela permet de tirer parti de la haute résolution de la RMN tout en extrayant simultanément des couplages dipolaires résiduels. Par conséquent, il est généralement souhaité que l'ordre induit sur la molécule soit relativement faible, typiquement inférieur à  $2 \times 10^{-3}$ . Pour atteindre un tel alignement,

interactions between the medium and solute have to be negligible. This cannot be predicted a priori and it is frequently necessary to empirically determine which medium is best suited for a particular case. If, however, binding to the medium occurs, line broadening is observed and the ordering will be too large to be of practical use. At present there is as yet not a single, universal alignment medium and it is generally advisable to test several for practical purposes.

The degree of alignment in any given liquid crystalline medium can easily be assessed by monitoring the deuterium quadrupolar splitting of the HDO signal in the sample prepared as an aqueous solution containing 90% H<sub>2</sub>O/10% D<sub>2</sub>O. This signal arises from exchange between bulk water and water bound to the oriented liquid crystal. Quadrupole splitting ranging from 10 Hz to 30 Hz are generally observed for the different media of variable composition. Interestingly, the degree of alignment of the solute macromolecule under investigation is not directly correlated to the deuterium quadrupole splitting across different liquid crystalline media, although a linear dependence on concentration is frequently observed for each individual medium. This is most likely due to the fact that the distribution and location of bound water molecules on the surfaces of the different large, anisotropic particles (lamellae of varying composition, phages and viruses, and assemblies of other surfactant phases) are very distinct and strongly dependent on the local surface structure. Nevertheless, measuring the deuterium quadrupole splitting of the HDO resonance is generally an excellent method to rapidly assess whether a medium is suitable for alignment purposes.

### 3. Alignment media

#### 3.1. Liquid crystalline media based on phospholipids

The first medium used for purposes of partial alignment of solute macromolecules with the magnetic field was a lyotropic liquid crystal consisting of binary mixtures of DMPC and DHPC. Binary mixtures of these phospholipids were initially believed to form disc shaped particles [8], commonly referred to as bicelles [9]. More recently, however, it was demonstrated that the morphology of these mixtures at concentrations and temperatures used for alignment in the magnetic field are stacked lamellar phases that align with their bilayer normal orthogonal to the field direction [10]. Using these dilute phospholipid mixtures, biological macromolecules such as proteins, nucleic acids or complexes thereof can be dissolved in the interstitial aqueous spaces between the layers, rendering their rotational

diffusion rates essentially unaffected by the lamellae. Residual dipolar couplings arising from the small degree of molecular alignment imparted onto the solute molecules by the oriented lamellae can therefore be measured. For most applications, the concentration of the phospholipids must be kept small (~ 5%) to avoid line broadening caused by unresolved dipolar couplings. However, under these dilute conditions the liquid crystal becomes less stable and the temperature range over which a stable nematic phase is observed is limited. In addition, samples may phase separate within the time required to collect the NMR data. Commonly used mixtures contain ~ 3% to 5% w/v DMPC/DHPC with a DMPC/DHPC ratio of ~ 3:1. They form stable liquid crystal phases over the temperature interval of 29–45 °C. This somewhat limited temperature range can be extended by using a ternary mixture of DHPC, DMPC and charged amphiphiles such as hexadecyl-(cetyl)trimethyl ammonium bromide (CTAB) or SDS [11] or by using shorter chain phospholipids instead of DMPC [12]. In addition, the long-term stability of the phospholipid based liquid crystal phase exhibits a strong pH dependence, restricting the solute pH to a narrow range around pH 7. Replacement of the diacyl phospholipids by non-hydrolyzable dialkyl analogs overcomes this limitation [13]. Despite all these improvements one has to keep in mind, that the stability of the liquid crystal phase is also affected by the solute, namely the protein solution under investigation. Solubility and stability of the protein naturally influence the choice of solvent conditions for the aqueous component, i.e. concentration, buffer choice and pH, and ionic strength. Each of these parameters affects the stability of the liquid crystalline phase in a complex manner.

#### 3.2. Nematic phases of rod-shaped viruses and filamentous phages

Suspensions of charged, rod-shaped viruses, such as tobacco mosaic virus (TMV) and filamentous bacteriophages fd/M13 and Pf1 were known to undergo a magnetic field induced isotropic-nematic phase transition at moderate concentrations [14, 15]. Solutions of magnetically aligned virus or phage therefore seemed like an attractive alternative to the above described phospholipid phases. Two independent reports of successfully using virus/phage solutions as a medium to measure residual dipolar couplings appeared, employing TMV and fd solutions [16] or Pf1 [17]. The molecular structures of the virus and phages are very similar. They are long, negatively charged rods, in which a cylinder of coat proteins is arranged in a helical fashion around either an RNA or DNA single stranded genome. TMV is approximately 15 nm wide and

~3000 nm long, while the bacteriophages are both ~6.6 nm in diameter and ~880 nm (fd/M13) or ~1900 nm (Pf1) long. Partial molecular alignment of the dissolved macromolecules arises from collisions with the aligned virus particles, thus imposing a preferred direction of diffusion rather than alignment via transient binding. It appears, that it is possible to align macromolecules over a wide range of temperatures and buffer conditions using dilute colloidal phage suspensions [18]. Indeed, the degree of alignment and the size of the residual dipolar coupling are related to the length of the phage, as expected from Onsager theory for semi-flexible charged rods [18]. Similar to findings for the phospholipid systems, high macroscopic viscosity due to the large viral particle size is observed. The microscopic tumbling rates of the dissolved macromolecules, however, should not be affected in the absence of binding to the particles. This can easily be established by measuring  $T_2$  relaxation times in the absence or presence of phage solution.

### 3.3. Liquid crystal phases of surfactants

The search for other, alternative and robust liquid crystalline media suitable for partially aligning biomolecules lead to investigations of dilute, quasi-ternary systems of surfactant/salt/alcohol, known to form Helfrich lamellar phases [19, 20]. These phases were thought to consist of bilayers, which can be swelled by solvent such that the spacing between the bilayers is much larger than the thickness of the bilayer itself. Thus, Helfrich lamellar phases were potentially another medium for studying partially aligned biological macromolecules. Prosser et al. used a 2% aqueous solution of CPCI/hexanol (1:1) in 200 mM NaCl and demonstrated that residual dipolar couplings up to 15 Hz could be measured on ubiquitin. Our laboratory investigated surfactant phases from CPBr/hexanol and NaBr and found that solutions of 3–6% CPBr/hexanol in 20–30 mM NaBr gave excellent results. Although initially expected to form lamellar phases, further characterization of the CPBr/hexanol/NaBr liquid crystal phase employed by us revealed, that the particle morphology consists of cylinders. Thus, the above quasi-ternary surfactant liquid crystalline phases are most likely cylindrical micelles [21].

Another lyotropic liquid crystalline phase suitable for alignment purposes is formed by alkyl-poly(ethylene glycol)/alcohol mixtures in water. These mixtures were known to form superstructures of stacked planar bilayers and the application of ~5% C12E5/hexanol and related phases for measuring residual dipolar couplings was demonstrated [22]. The advantage of these phases is their insensitivity to pH (as compared to phospho-

lipid phases), although, like with any thermotropic liquid crystal, only a limited temperature range is accessible. The major benefit of using these media is their low tendency to interact with proteins.

### 3.4. Other liquid crystals

In contrast to alignment caused by steric interactions with liquid crystalline media as described above, transient binding to oriented particles can also give rise to residual dipolar couplings. Such data has been reported for proteins interacting with purple membrane fragments [23, 24]. Binding is assumed to occur since significant line broadening and a large decrease in  $^{15}\text{N}$   $T_{1\rho}$  and  $T_2$  relaxation times of the solute proteins is observed. Obviously, here the degree of alignment depends on the binding properties of the biomolecule under investigation, and any information derived from residual couplings reflects the conformation in the bound state. The weak binding and fast exchange interaction between the medium and solute molecule is reminiscent of the transferred NOE effect.

The use of a suspension of cellulose crystallites for measuring residual dipolar couplings has also been reported [25]. This material can be prepared from wood pulp or filter paper by hydrolysis. Alignment in the magnetic field occurs due to the large negative diamagnetic anisotropy of individual cellulose crystallites. These particles generally have a length of several 100 nm and a width of ca 10 nm, and proteins can be dissolved in suspensions thereof.

Other media and methods for creating weakly aligned states for measuring residual dipolar couplings are still being searched for. It should be pointed out, at this juncture, that for all the media described above, sufficient care must be taken to ensure that the medium is not influencing the structural interpretation of the measured dipolar couplings. Flexible regions of proteins may transiently interact with the media and the bound conformation could become the dominant one, with other, non-binding conformations becoming under-represented. Typically, electrostatic and hydrophobic interactions between the solute proteins or nucleic acids will be weak enough, but this has to be verified for each individual case.

### 3.5. Strained gels

Recently, an alternative method for inducing weak alignment has been developed. This approach does not involve any liquid crystalline media, but rather exploits the anisotropy of strained polymeric gels. Either compression [26], or both vertical and radial squeezing of polyacrylamide gels [27], can be used. Compression of

a gel is achieved by using a susceptibility-matched plunger for pushing onto the gel. Stretching, in general, yields stronger and more uniform alignment. In this respect, stretching usually refers to radial compression accomplished by squeezing a larger diameter gel into a regular sample tube. Proteins are introduced into the gel matrix simply by diffusion. Such gels are extremely stable and inert and even allow the study of proteins under denaturing conditions [28]. In addition, gels may even be useful for aligning detergent solubilized systems that hitherto were not amenable to alignment. The principal disadvantage of using anisotropically compressed gels lies in a decrease in rotational diffusion rate of the dissolved macromolecules, especially for larger systems. Careful attention to the concentration of the gel as well as the degree of cross-linking has to be paid and adjusting these parameters to the system under investigation may be necessary.

### 3.6. The importance of multiple alignment media

There is no doubt that in the future more materials and methods will be discovered and exploited for imparting anisotropy onto biomolecules in order to extract residual dipolar couplings. There are several important reasons to have different alignment media available. First, not every medium is compatible with the properties of the molecules or systems under investigation. Proteins that interact with membranes are clearly not compatible with phospholipid-based media, and very flexible or partially folded proteins have a tendency to strongly interact with bicelles. Likewise, negatively charged molecules, such as nucleic acids, tend to bind to positively charged lamellae and, positively charged proteins can potentially interact with negatively charged phage particles at neutral pH values. This can result in an increase in the electrostatic component of the alignment, leading to large linewidths or to collapse of the liquid crystalline phase. As an example, the protein ubiquitin with a pI of ~ 6.5 interacts strongly with pf1, unless high ionic strength is used to screen the charges on the surface of the phage. Second, different alignment media frequently result in different orientations of the solute molecule with respect to the magnetic field, because the alignment tensors in two different alignment media will exhibit different orientations. This is an important property that allows lifting the degeneracies in the orientation of a given inter-atomic bond, inherent in the relationship between dipolar coupling and inter-nuclear vector orientation. A dipolar coupling measured in a given liquid crystalline medium positions the vector between the two coupled partners on one of the two possible, oppositely oriented cones. If the alignment tensor in the second medium

has a different orientation relative to the molecular frame of the molecule, the same vector will now reside on two different cones. Thus the true orientation of this particular inter-atomic vector will lie at the intersections between the two cones. Therefore, dipolar couplings measured in several independent media allow uniquely defining the associated vector orientations.

## 4. Structure refinement using residual dipolar couplings

A key aspect of any NMR structure determination is that the ensemble of calculated structures satisfies all of the experimental NMR constraints, exhibits only very small deviations from idealized covalent geometries, such as bond lengths, bond angles, and planarity, and displays good non-bonded contacts. It therefore is of utmost importance for devising any calculational strategy, that the global minimum of the target function is reliably and efficiently located. The use of residual dipolar couplings, which impose a tight restriction on the orientation of a bond (if measured for directly bonded nuclei) should therefore greatly improve the quality of traditional NMR structures, calculated based on NOE distance restraints, coupling constants and chemical shifts. The simplest way to incorporate the geometric content of the dipolar couplings into a structure calculation is by means of an error function. For each measured dipolar contribution a term

$$E_{\text{dipolar}} = k_{\text{dipolar}} (D_{\text{calc}} - D_{\text{obs}})^2 \quad (1)$$

(where  $k_{\text{dipolar}}$  is a force constant and  $D_{\text{calc}}$  and  $D_{\text{obs}}$  are the observed and calculated values of the dipolar couplings) is added to the various other terms, like distance and torsion angle restraints, covalent geometry and non-bonded contacts in the simulated annealing or minimizing protocols.  $E_{\text{dipolar}}$  is evaluated by calculating the  $\theta$  and  $\phi$  angles between the appropriate bond vectors (e.g., N–H, C $\alpha$ –H, C $\alpha$ –C', etc.) and an external arbitrary axis system. Since the orientation of the axis system is not known a priori, it is defined by an artificial tetra-atomic molecule comprising atoms X, Y, Z and O, with three mutually perpendicular bonds, X–O, Y–O, and Z–O, representing the  $x$ ,  $y$ , and  $z$  axes of the anisotropy tensor, which is allowed to float during the calculation [29, 30]. The expression for the residual dipolar coupling between two directly coupled nuclei can be simplified to the form:

$$D(\theta, \phi) = D_a \left( 3 \cos^2 \theta - 1 \right) + 3/2 R (\sin^2 \theta \cos 2 \phi) \quad (2)$$

where  $D_a$  and  $D_r$  are the axial and rhombic components of the traceless second rank diagonal tensor  $D$  (in Hz).

$R$  is the rhombicity defined by  $D_r/D_a$ ;  $\theta$  is the angle between the inter-atomic vector and the  $z$  axis of the tensor and  $\phi$  is the angle that describes the position of the projection of the inter-atomic vector on the  $x$ - $y$  plane of the tensor.  $D_a$  and  $D_r$  subsume a number of constants, including the gyromagnetic ratio of the two nuclei involved, the distance between them, the generalized order parameter  $S$  for internal motion of the inter-nuclear vector, the magnetic field strength and the medium permeability.

Naturally, using dipolar couplings in NMR structure determination and refinement is predicated on the assumption that motional averaging will not compromise the data. The magnitude of dipolar coupling depends on the generalized order parameter  $S$  for internal motions of the inter-atomic vector [31], thus different contributions have to be considered, at least in principle. Rather than using individual, residue specific  $S$  values, it seems reasonable to assume uniform  $S$  values for all those residues for which heteronuclear relaxation measurements indicate a well ordered conformation, as evidenced by experimental  $S^2$  values of 0.7–0.9 (corresponding to  $S$  values of 0.85 to 0.95). Dipolar coupling constraints for residues that experience either slow conformational exchange or low order parameters ( $S^2 < 0.6$ ) need to be excluded from the data set [29]. As an aside, it should be pointed out, that  $D_a$  and  $D_r$  scale with  $S$ , rather than  $S^2$ , thus the assumption of an overall  $S$  value introduces at most an error of a few percent in the dipolar couplings for the ordered regions of the molecule, well within the error of the experimental measurements.

In order to use equations (1) and (2) for structure refinement, the values of  $D_a$  and  $R$  have to be determined. They are obtained by either iteratively best fitting the alignment tensor for a given structure during the course of refinement as described above, or directly from the experimental data by examining the distribution of dipolar couplings [32]. A histogram of the ensemble of normalized residual dipolar couplings for the protein cyanovirin-N is illustrated in Fig. 1. It is possible to extract  $D_a$  and  $R$  from this distribution, given that different, fixed-distance inter-nuclear vector types in a molecule are approximately uniformly and isotropically distributed in space relative to the alignment tensor of the molecule. The magnitude of the axial and rhombic components of the molecular alignment tensor are related to the extrema and mode of the coupling histograms, which, in the absence of random errors, look almost like perfect powder patterns. The highest probability dipolar coupling value, therefore, coincides with the magnitude of the bond vector aligned along the  $x$  axis of the alignment tensor [32].

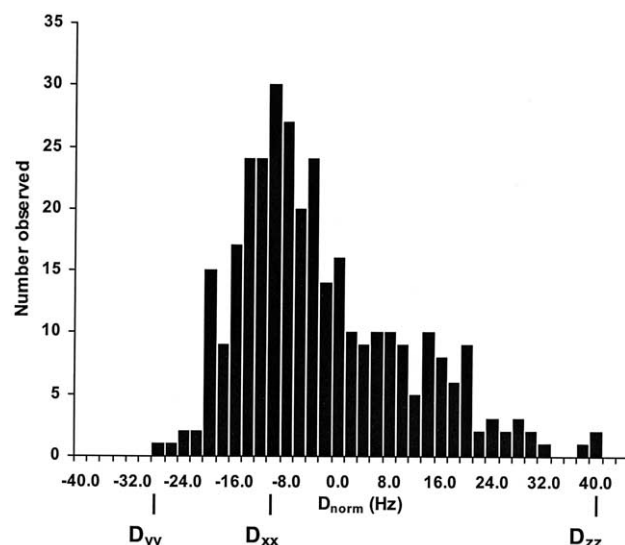


Fig. 1. Histogram depicting the distribution of normalized residual dipolar couplings measured for the protein CV-N. The mode ( $D_{xx}$ ) and extrema ( $D_{yy}$  and  $D_{zz}$ ) of the distribution are indicated.

Using this approach, the accuracy with which  $D_a$  and  $R$  are determined, clearly depends on the accuracy in the estimates for the two extrema and the maximum of the distribution, which in turn depends on the number of dipolar couplings observed and the degree of anisotropy in the orientation of the inter-nuclear vectors. In cases where it is not possible to measure a large set of dipolar couplings, one can also use a maximum likelihood method for extracting  $D_a$  and  $R$  [33]. Alternatively, singular value decomposition for calculation of the Saupe order matrix allows the determination of the axial and rhombic components of the alignment tensor [34]. Still another way to obtain the alignment tensor exists, if the alignment is purely steric. In this case, the alignment tensor can be predicted based on the shape of the solute molecule using an obstruction model [33]. This approach is very useful if an initial low-resolution structure is available, either based on a set of traditional NMR restraints or a model of a related molecule.

Refinement against dipolar couplings represents a difficult optimization problem, since each dipolar coupling is compatible with two orientations of the associated bond vectors, pointing in opposite directions. In addition, dipolar coupling constraints are of a very different qualitative nature, compared to NOE based distance constraints when used in a simulated annealing protocol. The success of the latter for structure determination based on distance constraints is based on the fact, that the experimentally determined inter-atomic distances are highly correlated. For example, if amino acid 20 is close to amino acid 90 within a folded protein structure, then the distance between atom X of amino

acid 20 and atom Y of amino acid 90 constrains the distance between X and any other atom of amino acid 90, such as atom Z. Furthermore, it is highly probable, that any atom on residue 21 is also close to one or more atoms of residue 90. In this manner, the potential surface of the optimization in three-dimensional space resembles a funnel with a relatively smooth surface. For the case of dipolar couplings, no such straightforward correlation exists. On the contrary, dipolar couplings (or bond-vector orientations) tend to compete with each other and with the distances. If, for example, during the simulated annealing a NH bond becomes oriented such, that its dipolar coupling is satisfied, this does not necessarily lead to a better agreement for the dipolar coupling of the adjacent NC' bond. Thus, each inter-atomic vector orientation represents an independent orientational parameter. As a result, calculational strategies based on dipolar couplings alone are fraught with difficulties. But even the simple addition of a constraint term for dipolar couplings to proven methodologies of NMR structure calculations can result in structures becoming trapped in deep local minima. As a consequence, the convergence properties of the procedure can be severely curtailed and careful adjustment of the protocols is necessary.

In order to avoid such trapping in the determination of a NMR structure, it is advantageous to employ a two stage simulated annealing protocol. In the first stage, all conventional experimental constraints, such as NOE and H-bond based distance constraints, dihedral angle constraints, coupling constants and carbon chemical shifts are employed to calculate an ensemble of models. During the second stage, each structure of the ensemble is refined against the dipolar couplings. These are added as constraints in an appropriate simulated annealing protocol, frequently employing both high temperature and low temperature slow cooling steps.

Inclusion of the additional dipolar constraints improves the precision of the structures considerably [35]. In particular, significant increases in coordinate precision are observed, both for backbone atoms and for side chains atoms. As an example, families of structures calculated without and with inclusion of residual dipolar couplings are displayed in Fig. 2 for the potent HIV-inactivating protein cyanovirin-N, vividly demonstrating the improved precision.

The most significant improvements in the quality of the structures upon inclusion of all residual dipolar couplings relate to the Ramachandran statistics. The percentage of residues found in the most allowed regions of the Ramachandran map generally increases from ca. 80% to over 90% and the number of bad contacts is reduced by more than 50%. Therefore,

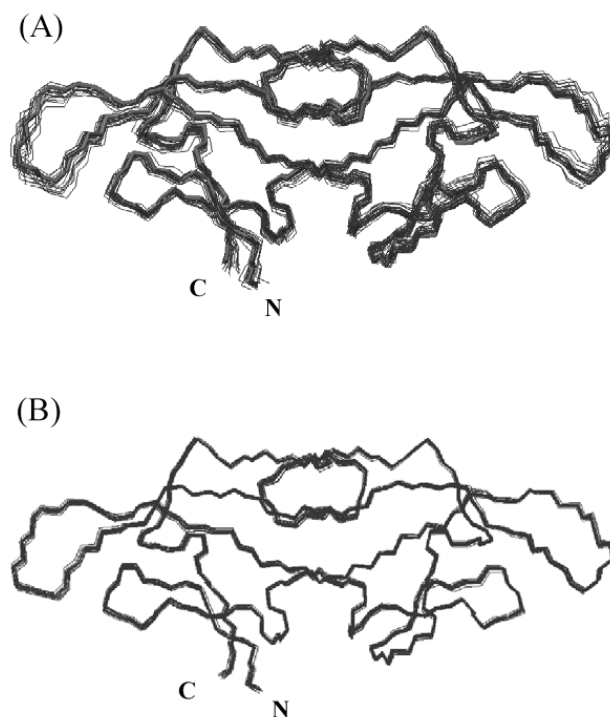


Fig. 2. Best-fit superpositions of backbone (N, C $\alpha$ , C') atoms of 20 CV-N structures calculated without (A) and with residual dipolar coupling constraints (B). Individual conformers are shown in blue and the minimized average structure in red.

structures calculated with a one-bond residual dipolar couplings exhibit superior packing characteristics, even without the inclusion of a conformational database potential [36].

## 5. Validation of protein folds

The most powerful and attractive use of residual dipolar couplings lies in their application for validation of structural models. In this context, structures may be derived from modeling, either ab initio or homology, or from low-resolution experimental data. If the model structure is accurate, one will observe excellent agreement between the dipolar couplings calculated based on the structure and the experimentally measured ones. Exploitation of this fact may lead to the most direct and important contribution of NMR based methodology in 'Structural Genomics'. Although large efforts are underway to determine as many protein structures by NMR in a high throughput manner, traditional methodology for NMR structure determination for proteins of intermediate size (~30–50 kDa) is still relatively slow. For instance, to solve the structure of the 40 kDa complex between the N-terminal domain of enzyme I and HPr, the NMR measurement time alone extended over almost

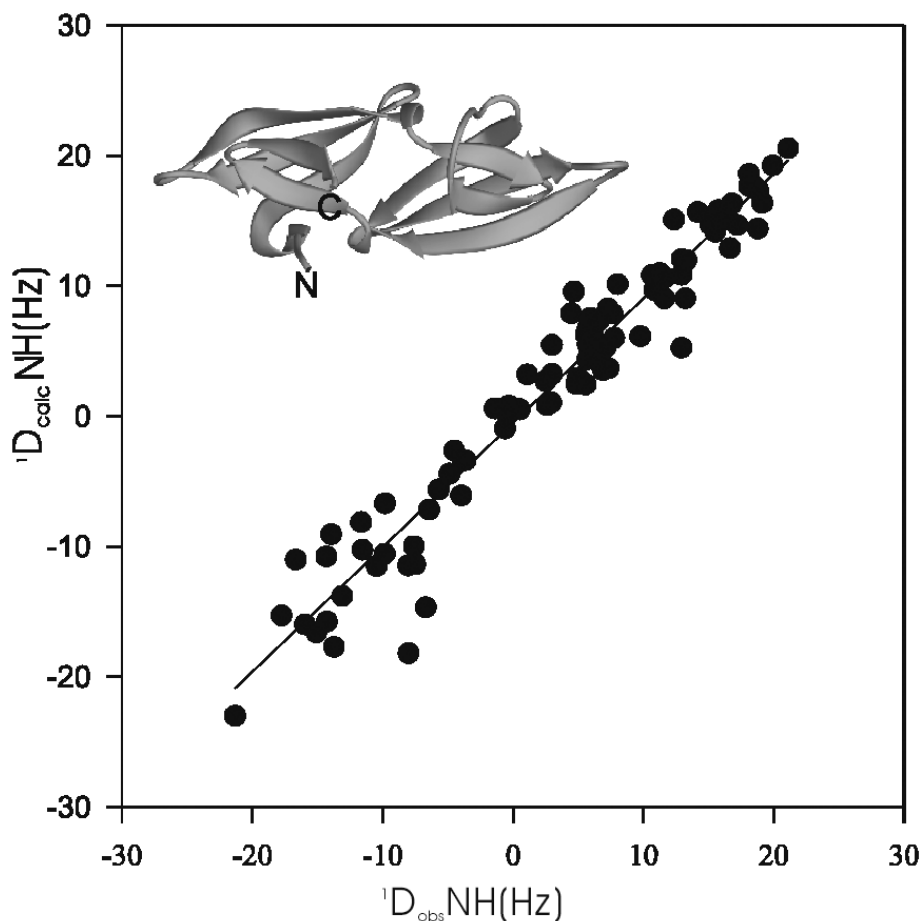


Fig. 3. Correlation between experimentally measured dipolar couplings ( $^1D_{\text{obs}}\text{NH}$ ) and those calculated ( $^1D_{\text{calc}}\text{NH}$ ) on the basis of the coordinates of a homology model (inset) for a circular permuted variant of the protein CV-N.

five months [37], with an additional time of at least 6 months for data interpretation and structure calculation. It therefore seems imperative to explore alternative avenues, if NMR methodology to become more powerful for structure determination in the post-genomic era. Exploitation of strategies based on residual dipolar couplings may overcome the traditional shortcomings indicated above, and could lead to new conceptual applications of NMR. Here again, the alignment tensor or Saupe matrix has to be determined and any of the methodologies outlined previously for refinement can be applied for this purpose. Naturally, it is important to include sufficient experimentally determined couplings. Given that there are five independent Saupe matrix elements, it generally is easy to include a much larger number of observables (measured residual dipolar couplings) than variables in the fitting procedure, rendering the SVD approach superior to the shape-based prediction of the alignment tensor.

A very promising direction combines structure prediction and experimental validation or selection. Based on our increasing understanding of the important fac-

tors that govern folding and stability of proteins, it has been proposed that only a finite set of protein folds exist [38]. With improving force fields it may be conceivable, that for a particular amino acid sequence a set of several thousand different possible folds can be generated, one of which may be the correct one. The crucial question in this scenario then remains, how to identify this particular one. Using NMR may provide the answer. It is relatively fast and straightforward to obtain backbone resonance assignments for isotopically labeled proteins, and the measurement of residual dipolar couplings in alignment media can also be achieved with relative ease. It therefore is possible to experimentally obtain residual dipolar couplings for the backbone, which in turn can be compared with those calculated for the different theoretically predicted structures. This approach is illustrated with an example in Fig. 3. A homology model for a circular permuted variant of cyanovirin-N was constructed and, based on the coordinates of this model, residual dipolar couplings were calculated. Comparison of the calculated and experimentally measured couplings allowed veri-



fication of the modeled fold. Thus, NMR can be used as an experimental filter for any theoretically predicted structures. Initial results exploring such methodologies are clearly promising.

## 6. Concluding remarks

The accessibility of anisotropic NMR parameters in solution for biological macromolecules has opened the door for future imaginative exploitation of this diverse wealth of physical information. Applications with respect to improving the accuracy of protein structures [39] in defining the long-range orientation of the RNA in a protein–RNA complex [40] and domain orientation

in multi-domain proteins [41–44], as well as recognition of protein folds [45, 46] have already been reported. No doubt further developments will occur. In the era of structural genomics, NMR is poised to make a major contribution. Since dipolar constraints can be readily measured on partially aligned proteins, they can be incorporated into powerful methodologies for validation of theoretical models. Not every protein structure of the completed human genome and those of model organisms will be solved experimentally by NMR or X-ray crystallography. However, all of them are amenable to structure prediction, and validation of these predicted structures and folds can be achieved rapidly by NMR using residual dipolar couplings.

**Acknowledgements.** I am indebted to all my collaborators and colleagues mentioned in the text and references who have developed the technologies described in this article and who have been a constant source of stimulating discussions. Particularly, Drs. Ad Bax, Nico Tjandra and Anatoliy Dobrodumov provided invaluable intellectual contributions. The work in the author's laboratory was in part supported by the Intramural AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health.

## References

- [1] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
- [2] G.M. Clore, A.M. Gronenborn, Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy, *CRC Crit. Rev. Biochem. Mol. Biol.* 24 (1989) 479–564.
- [3] N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium [erratum in *Science* (1997) 278, 1697], *Science* 278 (1997) 1111–1114.
- [4] J.R. Tolman, J.H. Prestegard, A quantitative  $J$ -correlation experiment for the accurate measurement of one-bond amide  $^{15}\text{N}$ – $^1\text{H}$  couplings in proteins, *J. Magn. Reson. Ser. B* 112 (1996) 245–252.
- [5] N. Tjandra, S. Grzesiek, A. Bax, Magnetic field dependence of nitrogen–proton  $J$  splittings in  $^{15}\text{N}$ -enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling, *J. Am. Chem. Soc.* 118 (1996) 6264–6272.
- [6] J.H. Prestegard, New techniques in structural NMR – anisotropic interactions, *Nat. Struct. Biol.* 5 (Suppl) (1998) 517–522.
- [7] A. Bax, G. Kontaxis, N. Tjandra, Dipolar couplings in macromolecular structure determination, *Methods Enzymol.* 339 (2001) 127–174.
- [8] C.R. Sanders II, J.P. Schwonek, Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR, *Biochemistry* 31 (1992) 8898–8905.
- [9] C.R. Sanders II, G.C. Landis, Reconstitution of membrane proteins into lipid-rich bilayered mixed micelles for NMR studies, *Biochemistry* 34 (1995) 4030–4040.
- [10] M-P. Nieh, C.J. Glinka, S. Krueger, R.S. Prosser, J. Katsaras, SANS study on the effect of lanthanide ions and charged lipids on the morphology of phospholipid mixtures, *Biophys. J.* 82 (2002) 2487–2498.
- [11] J.A. Losonczi, J.H. Prestegard, Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules, *J. Biomol. NMR* 12 (1998) 447–451.
- [12] H. Wang, M. Eberstadt, E.T. Olejniczak, R.P. Meadows, S.W. Fesik, A liquid crystalline medium for measuring residual dipolar couplings over a wide range of temperatures, *J. Biomol. NMR* 12 (1998) 443–446.
- [13] M. Ottiger, A. Bax, Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values, *J. Biomol. NMR* 13 (1999) 187–191.
- [14] H. Nakamura, K. Okano, Pretransitional phenomena in the isotropic phase of a lyotropic liquid crystal of bacterial virus fd, *Phys. Rev. Lett.* 50 (1983) 186–189.
- [15] S. Fraden, G. Maret, D.L.D. Caspar, R.B. Meyer, Isotropic–nematic phase transition and angular correlations in isotropic suspensions of tobacco mosaic virus, *Phys. Rev. Lett.* 63 (1989) 2068–2071.
- [16] G.M. Clore, M.R. Starich, A.M. Gronenborn, 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 (1998) 10571–10572.
- [17] M.R. Hansen, L. Mueller, A. Pardi, Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions, *Nat. Struct. Biol.* 5 (1998) 1065–1074.
- [18] L.G. Barrientos, J.M. Louis, A.M. Gronenborn, Characterization of the cholesteric phase of filamentous bacteriophage fd for molecular alignment, *J. Magn. Reson.* 148 (2001) 159–162.
- [19] R.S. Prosser, J.A. Losonczi, I.V. Shiyonovskaya, Use of a novel aqueous liquid crystalline medium for high-resolution NMR of macromolecules in solution, *J. Am. Chem. Soc.* 120 (1998) 11010–11011.
- [20] L.G. Barrientos, C. Dolan, A.M. Gronenborn, Characterization of surfactant liquid crystal phases suitable for molecular alignment and measurement of dipolar couplings, *J. Biomol. NMR* 16 (2000) 329–337.
- [21] L.G. Barrientos, K. Gawrisch, N. Cheng, A.C. Steven, A.M. Gronenborn, Structural Characterization of the dilute aqueous surfactant solution of cetylpyridinium bromide/hexanol/sodium bromide, *Langmuir* 18 (2002) 3773–3779.
- [22] M. Rückert, G. Otting, Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments, *J. Am. Chem. Soc.* 122 (2000) 7793–7797.
- [23] B.W. Koenig, J.S. Hu, M. Ottiger, S. Bose, R.W. Hendler, A. Bax, NMR measurement of dipolar couplings in proteins aligned by transient binding to purple membrane fragments, *J. Am. Chem. Soc.* 121 (1999) 1385–1386.
- [24] H.J. Sass, F. Cordier, A. Hoffmann, M. Rogowski, A. Cousin, J.G. Omichinski, H. Lowen, S. Grzesiek, Purple membrane induced alignment of biological macromolecules in the magnetic field, *J. Am. Chem. Soc.* 121 (1999) 2047–2055.
- [25] K. Fleming, D. Gray, S. Prasanna, S. Matthews, Cellulose crystallites: a new and robust liquid crystalline medium for the measurement of residual dipolar couplings, *J. Am. Chem. Soc.* 122 (2000) 5224–5225.
- [26] R. Tycko, F.J. Blanco, Y. Ishii, Alignment of biopolymers in strained gels: a new way to create detectable dipole–dipole couplings in high resolution biomolecular NMR, *J. Am. Chem. Soc.* 122 (2000) 9340–9341.

- [27] H.J. Sass, G. Musco, S.J. Stahl, P.T. Wingfield, S. Grzesiek, 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 (2000) 303–309.
- [28] D. Shortle, M.S. Ackerman, Persistence of native-like topology in a denatured protein in 8 M urea, *Science* 293 (2001) 487–489.
- [29] N. Tjandra, J.G. Omichinski, A.M. Gronenborn, G.M. Clore, A. Bax, 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 (1997) 732–738.
- [30] G.M. Clore, A.M. Gronenborn, N. Tjandra, Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude, *J. Magn. Reson.* 131 (1998) 159–162.
- [31] G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. I. Theory and range of validity, *J. Am. Chem. Soc.* 104 (1982) 4546–4559.
- [32] G.M. Clore, A.M. Gronenborn, A. Bax, 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 (1998) 216–221.
- [33] J.J. Warren, P.B. Moore, A maximum likelihood method for determining  $D_a^{PQ}$  and  $R$  for sets of dipolar coupling data, *J. Magn. Reson.* 149 (2001) 271–275.
- [34] J.A. Losonczi, M. Andrec, M.W. Fischer, J.H. Prestegard, Order matrix analysis of residual dipolar couplings using singular value decomposition, *J. Magn. Reson.* 138 (1999) 334–342.
- [35] G.M. Clore, A.M. Gronenborn, New methods of structure refinement for macromolecular structure determination by NMR, *Proc. Natl Acad. Sci. USA* 95 (1998) 5891–5898.
- [36] J. Kuszewski, A.M. Gronenborn, G.M. Clore, Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases, *Protein Sci.* 5 (1996) 1067–1080.
- [37] D.S. Garrett, Y.J. Seok, A. Peterkofsky, A.M. Gronenborn, G.M. Clore, Solution structure of the 40 000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr, *Nat. Struct. Biol.* 6 (1999) 166–173.
- [38] Y.I. Wolf, N.V. Grishin, E.V. Koonin, Estimating the number of protein folds and families from complete genome data, *J. Mol. Biol.* 299 (2000) 897–905.
- [39] M. Zweckstetter, A. Bax, Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR, *J. Am. Chem. Soc.* 122 (2000) 3791–3792.
- [40] P. Bayer, L. Varani, G. Varani, Refinement of the structure of protein-RNA complexes by residual coupling analysis, *J. Biomol. NMR* 14 (1999) 149–155.
- [41] M.W.F. Fischer, J.A. Losonczi, J.L. Weaver, J.H. Prestegard, Domain orientation and dynamics in multidomain proteins from residual dipolar couplings, *Biochemistry* 38 (1999) 9013–9022.
- [42] N.R. Skrynnikov, N.K. Goto, D. Yang, W.Y. Choy, J.R. Tolman, G.A. Mueller, L.E. Kay, 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 (2000) 1265–1273.
- [43] J.J. Chou, S. Li, A. Bax, Study of conformational rearrangement and refinement of structural homology models by the use of heteronuclear dipolar couplings, *J. Biomol. NMR* 18 (2000) 217–227.
- [44] L.G. Barrientos, J.M. Louis, I. Botos, T. Mori, Z. Han, B.R. O’Keefe, M.R. Boyd, A. Wlodawer, A.M. Gronenborn, The domain-swapped dimer of Cyanovirin-N is in a metastable folded state. Reconciliation of X-ray and NMR structures, *Structure* 10 (2002) 673–686.
- [45] A. Annala, H. Aittio, E. Thulin, T. Drakenberg, Recognition of protein folds via dipolar couplings, *J. Biomol. NMR* 14 (1999) 223–230.
- [46] F. Delaglio, G. Kontaxis, A. Bax, Protein structure determination using molecular fragment replacement and NMR dipolar couplings, *J. Am. Chem. Soc.* 122 (2000) 2142–2143.