Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 162 (2003) 385–395
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Internal consistency of NMR data obtained in partially aligned biomacromolecules

Lukáš Žídek, Petr Padrta, Josef Chmelík Jr., and Vladimír Sklenář*

National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kotlářská 2, 61137 Brno, Czech Republic

Received 24 October 2002; revised 19 February 2003

Abstract

A method of testing structure-related NMR data prior to structure calculations is presented. The test is applicable to second rank tensor interactions (dipolar coupling, anisotropic chemical shielding, and quadrupolar interaction) observed in partially aligned samples of biomacromolecules. The method utilizes the fact that only limited number of frequencies corresponding to the mentioned interactions can be measured independently in a rigid fragment of the macromolecule. Additional values can be predicted as linear combinations of the set of independent frequencies. Internal consistency of sufficiently large sets of frequencies measured in individual molecular fragments is tested by comparing the experimental data with their predicted values. The method requires only knowledge of local geometry (i.e., definition of the interaction tensors in the local coordinate frames of the fragments). No information about the alignment or shape of the molecule is required. The test is best suited for planar fragments. Application to peptide bonds and nucleic acid bases is demonstrated.

Keywords: NMR; Residual dipolar couplings; Chemical shift anisotropy; Peptide plane; Nucleic acid base

1. Introduction

Second rank tensor interactions measured in partially aligned samples greatly improved structure determination of large biologically relevant molecules, such as proteins and nucleic acids [1,2,5,9,15–17,19]. Frequencies corresponding to these interactions depend on the averaged orientations of the interacting nuclei with respect to the external magnetic field. Useful information about relative orientations of submolecular fragments is thus provided regardless of the through-bond or spatial distance of the observed nuclei. Quantitative description of the observed interactions can be expressed in terms of two second rank tensors, one describing the spatial dependence of the interactions and the other defining averaged alignment of the studied molecule [10]. The process of structure determination can be viewed as a search for a matrix representing the first, geometry-related, tensor in a single coordinate frame. The second tensor is represented by the Saupe order matrix [8].

Two different approaches to incorporation of the second rank tensor interactions have been proposed and applied [15]. The first approach is based on the optimization of spatial coordinates defining the geometry-related matrix for the whole molecule. As pointed out by Moltke and Grzesiek [10], a priori knowledge of the order matrix is not required because such optimization is a linear least squares problem with respect to the elements of the order matrix. Nevertheless, principal values of the order matrix are usually determined prior to structure refinement from the distribution of the measured data [3].

The other approach is based on decomposing the studied macromolecule into smaller fragments of known local geometry. Five independent elements of the Saupe order matrix can be determined in a coordinate frame of an individual fragment if at least five frequencies (e.g., dipolar coupling constants) are measured within the fragment. As a result, a set of order matrices is obtained, each expressed in local coordinate frame of the corre-

* Corresponding author. Fax: +420-541-129-506.
E-mail address: sklenar@chemi.muni.cz (V. Sklenář).

1090-7807/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1090-7807(03)00116-2

sponding fragment. Rotational matrices transforming the individual coordinate systems into a common frame are then searched for. Euler angles as arguments in the individual rotational matrices define relative orientations of the individual fragments. The described procedure represents a straightforward application of linear algebra not requiring an elaborate structure refinement protocol. However, it should be noted that several solutions are obtained for each frequency set unless data acquired in different aligning media are available. This orientational degeneracy is a major limitation of using the second rank tensor interactions in structure determination.

The latter approach can be applied to relatively large fragments of size of protein domains or subunits [8]. The large size of the fragments allows to exploit the most easily measurable frequencies. For example, a large number of one-bond ^{1}H–^{13}C and ^{1}H–^{15}N residual dipolar couplings can be measured with high precision and relatively uniform spatial distribution. On the other hand, geometry of the large fragments, usually obtained from X-ray diffraction data, might not be sufficiently precise and/or accurate for the described procedure [8].

The uncertainty in the structure of individual fragments may be eliminated by choosing small rigid fragments of size of the peptide plane or the nucleic acid bases [7,11]. Hus et al. [7] recently demonstrated the power of this approach in a case when comparison of the data measured in differently aligned phases helped to remove the degeneracy of the determined Euler angles. They showed that a protein backbone conformation can be determined solely from the dipolar coupling data acquired for individual peptide planes and tetrahedral junctions of the α carbons.

However, the described use of very small fragments presents certain difficulties. Reduction of the size of the fragments also results in reduction of the number of measurable frequencies. In the most common case of using dipolar couplings, the structure determination requires measurement of two-bond coupling constants and one-bond coupling constants of nuclear pairs not involving protons, in addition to the easily measurable one-bond ^{1}H–^{13}C and ^{1}H–^{15}N couplings. Values of the mentioned additional couplings (and often of the corresponding scalar couplings too) are much lower than linewidths of the peaks in NMR spectra. Using small fragments is thus more experimentally demanding. Several pulse sequences were designed in order to measure small couplings in proteins and nucleic acids reliably [14,18,20,21]. Nevertheless, the small couplings are much more sensitive to experimental errors. Therefore, the quality of the experimental data should be carefully checked and only the reliable values used in structure calculation.

As noted by Losonczi et al. [8], determination of the order matrix by singular value decomposition also allows the identification of the measured values of frequencies that are not consistent with the remaining data. Unreliable data can be identified by comparison with the corresponding values back-calculated using the determined order matrix. This study presents a method for identification of unreliable data *prior* to structural analysis. The method is completely independent of the order matrix and can be applied even in cases when the complete order matrix cannot be determined.

2. Theory

Contributions of the λth second rank tensor interaction to the nuclear Hamiltonian can be expressed in the general form

$$H_\lambda = g_\lambda \sum_{k=x,y,z} \sum_{l=x,y,z} T_{\lambda,kl} A_{\lambda 1,k} A_{\lambda 2,l}, \tag{1}$$

where the vectors $\vec{A}_{\lambda 1}$ and $\vec{A}_{\lambda 2}$ represent nuclear spin operators or external magnetic field, T_λ is the second rank tensor describing the spatial dependence of the λth interaction, and g_λ is the relevant interaction constant.

In the high-field and weak-coupling approximation, contributions in the directions of the x and y axes of the laboratory coordinate system are neglected. As the tensor T_λ is defined in a coordinate frame associated with the local geometry of a particular molecular fragment, the average Hamiltonian of the λth interaction can be expressed as

$$\langle H_\lambda \rangle = g_\lambda A_{\lambda 1,z} A_{\lambda 2,z} \sum_{m=X,Y,Z} \sum_{n=X,Y,Z} S_{mn} T_{\lambda,mn}, \tag{2}$$

where the Saupe order matrix S represents the time-averaged transformation from the local coordinate frame X, Y, Z, associated with the molecule undergoing a rotational motion, to the laboratory coordinate frame x, y, z, given by the static magnetic field.

The second rank tensor interactions can be observed as resonance frequencies of transitions between spin states $|1\rangle$ and $|2\rangle$.

$$\omega_\lambda = c_\lambda \sum_{m=X,Y,Z} \sum_{n=X,Y,Z} S_{mn} T_{\lambda,mn}, \tag{3}$$

where

$$c_\lambda = \frac{g_\lambda}{\hbar} (\langle 2|A_{\lambda 1,z} A_{\lambda 2,z}|2\rangle - \langle 1|A_{\lambda 1,z} A_{\lambda 2,z}|1\rangle). \tag{4}$$

Five independent elements of the order matrix can be treated as vector elements S_j and Eq. (3) can be therefore written as

$$\frac{\omega_\lambda}{c_\lambda} = \sum_{j=1}^{5} S_j T_{\lambda j}. \tag{5}$$

Fundamental geometric relations limit the number of frequencies ω_λ that can be measured independently in a

given fragment. Any frequency ω_μ measured in addition to the independent set of L frequencies ω_λ ($\lambda = 1, \ldots, L$) may thus be expressed as a linear combination of L independent frequencies ω_λ

$$\frac{\omega_\mu}{c_\mu} = \sum_{\lambda=1}^{L} b_{\mu\lambda} \frac{\omega_\lambda}{c_\lambda}. \tag{6}$$

Substituting Eq. (5) into Eq. (6) shows that the coefficients $b_{\mu\lambda}$ can be obtained by solving the matrix equation

$$T_{\mu j} = \sum_{\lambda=1}^{L} T_{\lambda j} b_{\mu\lambda}. \tag{7}$$

If the number of measured frequencies M is higher than the number of independent frequencies L, internal consistency of the measured frequencies can be tested by checking whether the measured values correspond to the expected linear combinations, as described by Eq. (6).

In the case of a general, three-dimensional, spatial orientation of the principal axes of T, two ways of testing the internal consistency of the measured frequencies are possible. The first approach utilizes the fact that all five vector elements S_j can be determined if five independent frequencies ω_λ ($L = 5$) are measured within the rigid fragment of a known geometry [8] and if the matrix elements $T_{\lambda j}$ are known. Once the Saupe order matrix elements are known, additional frequencies ω_μ can be predicted and compared to the experimental values using Eq. (5). The other possibility is to use Eq. (6) directly, without evaluating the vector elements S_j.

The situation is different in special cases, when the number of independent frequencies L is reduced. For example, only three independent frequencies ($L = 3$) can be determined in the frequent case of residual dipolar couplings measured in a planar molecular fragments. Note that five Saupe order matrix elements S_j are still needed to describe the time-averaged orientation of the planar fragment in space. Since all five vector elements S_j cannot be obtained from Eq. (5), the Saupe matrix cannot be used to predict additional frequencies in the planar fragments. However, the direct use of Eq. (6) is still possible if more than three frequencies are measured ($L > 3$). It can be thus concluded that the direct prediction of the measured frequencies, based on Eqs. 6 and 7, is a general method of testing the internal consistency of the measured frequencies, particularly important in the case of the planar molecular fragments. Application of the proposed method to the case of planar fragment is treated in the following text.

Natural choice of the local coordinate system in the mentioned case of planar fragments is such that the Z axis is perpendicular to the plane. The λth row of the matrix T is then given by the elements

$$\begin{aligned}
T_{\lambda 1} &= T_{\lambda,XX} - T_{\lambda,ZZ} = 3\cos^2\phi_\lambda,\\
T_{\lambda 2} &= T_{\lambda,YY} - T_{\lambda,ZZ} = 3\sin^2\phi_\lambda,\\
T_{\lambda 3} &= T_{\lambda,XY} = -3\sin\phi_\lambda\cos\phi_\lambda,\\
T_{\lambda 4} &= T_{\lambda,XZ} = 0,\\
T_{\lambda 5} &= T_{\lambda,YZ} = 0,
\end{aligned} \tag{8}$$

where ϕ_λ is the the angle between the X axis and the internuclear vector corresponding to the λth dipolar coupling. The same definition applies to $T_{\lambda j}$ in the case of axially-symmetric anisotropic component of the chemical shielding tensor σ with its principal Y' axis perpendicular to the plane, where ϕ_λ is the the angle between X axis and the principal Z' axis. The asymmetry parameter η_λ, defined by the principal values $|\sigma_{\lambda,Z'Z'}| \geqslant |\sigma_{\lambda,Y'Y'}| \geqslant |\sigma_{\lambda,X'X'}|$ as

$$\eta_\lambda = \frac{\sigma_{\lambda,X'X'} - \sigma_{\lambda,Y'Y'}}{\sigma_{\lambda,Z'Z'}}, \tag{9}$$

has to be introduced in the case of asymmetric σ. The elements $T_{\lambda 1}$, $T_{\lambda 2}$, and $T_{\lambda 3}$ are then defined as

$$\begin{aligned}
T_{\lambda 1} &= 3\cos^2\phi_\lambda + \eta_\lambda(1 + \sin^2\phi_\lambda),\\
T_{\lambda 2} &= 3\sin^2\phi_\lambda + \eta_\lambda(1 + \cos^2\phi_\lambda),\\
T_{\lambda 3} &= -(3 - \eta_\lambda)\sin\phi_\lambda\cos\phi_\lambda
\end{aligned} \tag{10}$$

for Y' perpendicular to the plane of the fragment, as

$$\begin{aligned}
T_{\lambda 1} &= 3\cos^2\phi_\lambda - \eta_\lambda(1 + \sin^2\phi_\lambda),\\
T_{\lambda 2} &= 3\sin^2\phi_\lambda - \eta_\lambda(1 + \cos^2\phi_\lambda),\\
T_{\lambda 3} &= -(3 + \eta_\lambda)\sin\phi_\lambda\cos\phi_\lambda
\end{aligned} \tag{11}$$

for X' perpendicular to the plane of the fragment, or as

$$\begin{aligned}
T_{\lambda 1} &= -3 - \eta_\lambda(\cos^2\phi_\lambda - \sin^2\phi_\lambda),\\
T_{\lambda 2} &= -3 + \eta_\lambda(\cos^2\phi_\lambda - \sin^2\phi_\lambda),\\
T_{\lambda 3} &= 2\eta_\lambda\sin\phi_\lambda\cos\phi_\lambda
\end{aligned} \tag{12}$$

for Z' perpendicular to the plane of the fragment, where ϕ_λ is the angle between the X and Y' axes.

3. Methods

The proposed method of testing internal consistency of the measured data can be described in three steps. In step 1, the measured frequencies are normalized with respect to c_λ. The numerical values of c_λ for dipolar (c_λ^{D}), anisotropic chemical shift (c_λ^{S}), and quadrupolar contributions (c_λ^{Q}) are given as [10]

$$c_\lambda^{\mathrm{D}} = -\frac{\mu_0}{8\pi}\frac{\gamma_1\gamma_2\hbar}{r_{12}^3}, \tag{13}$$

$$c_\lambda^{\mathrm{S}} = \frac{1}{3}\gamma_1 B_0 \Delta\sigma, \tag{14}$$

$$c_\lambda^{\mathrm{Q}} = \frac{-eQ}{8\hbar I(2I-1)} V_{Z'Z'}, \tag{15}$$

where γ_1 and γ_2 are the magnetogyric ratios of the interacting nuclei, r_{12} is their distance, B_0 is the external magnetic field induction, $\Delta\sigma$ is the chemical shift anisotropy, eQ is the nuclear quadrupolar moment, I is the nuclear spin quantum number, $V_{Z'Z'}$ is the expectation value of the electric field intensity gradient at the position of the nucleus, μ_0 is the magnetic permeability of vacuum, and $\hbar$ is the Planck constant divided by 2π. In step 2, L measured normalized frequencies Ω_λ are selected as a basis set. In step 3, expected values of the remaining $M-L$ normalized frequencies are calculated according to Eq. (6). The coefficients $b_{\mu\lambda}$ are obtained by solving Eq. (7).

When the set of expected frequencies calculated using the selected basis is obtained, the algorithm returns to step 2. A new basis set of L frequencies is selected and steps 2 and 3 are repeated for all combinations of measured frequencies defining all possible basis sets. The described algorithm thus results in an array of $M!/L!(M-L)!$ sets of calculated frequencies. However, it should be emphasized that certain combinations of experimental values may form basis sets that give errors in the calculated frequencies much higher than errors in the experimental values used as the basis sets. Such combinations can be easily identified and discarded, as described in Section 5. The number of used basis sets is then lower than the number of possible combinations.

The result can be displayed graphically by plotting the calculated expected frequencies as a function of the selected basis sets. If all data are consistent, all calculated expected values of the μth frequency $\Omega_\mu^{\mathrm{calc}}$ are found along a horizontal line drawn at the measured value of $\Omega_\mu^{\mathrm{exp}}$. If one measured value $\Omega_\zeta^{\mathrm{exp}}$ is inconsistent with the rest of experimental data, expected values of frequencies $\Omega_{\mu\neq\zeta}^{\mathrm{calc}}$, calculated from basis sets containing the erroneous measured value $\Omega_\zeta^{\mathrm{exp}}$, deviate from the horizontal lines, and the horizontal line of calculated expected $\Omega_\zeta^{\mathrm{calc}}$ is shifted from the measured value of $\Omega_\zeta^{\mathrm{exp}}$.

4. Results

Application of the method is first demonstrated on an example of planes of nucleic acid bases. The experimental data has been obtained in our laboratory for a DNA minihairpin of the sequence d(GCGA AGC), partially aligned using 20-mg/ml solution of bacteriophage Pf1 [13,18]. In order to keep numerical values of the normalized frequencies Ω_λ in a convenient range, data are expressed as residual dipolar coupling constants (normalized to one type of coupling) in this study. For example, four dipolar couplings measured in guanine 3, $^1\mathrm{D(C8H8)} = (0.2 \pm 0.8)$ Hz, $^2\mathrm{D(N7H8)} = (0.1 \pm 0.2)$ Hz, $^1\mathrm{D(N9C8)} = (0.6 \pm 0.2)$ Hz, and $^2\mathrm{D(N9H8)} = (0.3 \pm 0.2)$ Hz, were normalized with respect to ^{1}D(C8H8), obtaining the values of (0.2 ± 0.8) Hz, (-1.5 ± 3.8) Hz, (-12.4 ± 6.2) Hz, and (-5.8 ± -3.8) Hz, respectively. Fig. 1 shows the result of testing the listed residual dipolar couplings in guanine 3. The plot in Fig. 1 contains only one calculated expected coupling for each of the four basis sets. Since the calculated values lay close to the lines representing the measured values, the plot validates the internal consistency of the measured data. The internal consistency can be also expressed quantitatively as the root mean square deviation of the calculated couplings from their experimental values (RMSD(Ω)). In the discussed case, RMSD(Ω) of 0.74 Hz was obtained. The fact that the achieved RMSD(Ω) was lower than the normalized estimated experimental errors, as listed above, indicated a very good internal consistency of the residual dipolar couplings measured in guanine 3. A set of four measured couplings ($M = 4$) represents the minimal number of experimental data for which the internal consistency can be checked. In this minimal case, the method only shows whether the measured data are internally consistent or not.

When more than four couplings are measured, it is also possible to reveal which value is inconsistent with the remaining data. Such identification of inconsistent values is demonstrated in Fig. 2. Panel A contains the

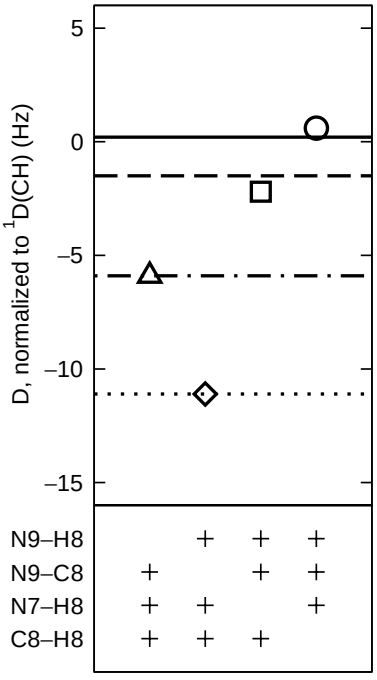

Fig. 1. Comparison of calculated expected values of residual dipolar couplings in guanine 3 of the DNA hairpin (symbols) with their measured values (horizontal lines): ^{1}D(C8H8), circle and solid line, ^{2}D(N7H8), square and dashed line, ^{1}D(N9C8), diamond and dotted line, and ^{2}D(N9H8), triangle and dash-and-dot line. Basis sets used for calculation of individual expected couplings are indicated below the plot. All couplings were normalized with respect to ^{1}D(C8H8).

graphical output of testing seven residual dipolar couplings measured in adenine 4. Four calculated values are plotted for each basis set of three experimental values that are indicated with the color-coded plus signs bellow the plot. Note that basis sets containing at the same time couplings ^{1}D(C8H8) and ^{2}D(N3H2) were omitted since they correspond to nearly collinear vectors (the angle between these two vectors was equal to 0.3° in the adenine geometry used in this study). The number of frequency sets is thus reduced from 35 to 25. Careful inspection of the plot reveals that poor prediction of the measured couplings was obtained when the basis set contained ^{2}D(N7H8) and/or ^{2}D(N9H8), indicated with orange and cyan in the graph, respectively. In accordance with this finding, the predicted values of the mentioned couplings systematically deviated from their measured values. The described analysis thus clearly identifies two measured couplings (between nuclei N7–H8 and N9–H8) that are inconsistent with the remaining data. When the inconsistent couplings were excluded from the tested set, all calculated values fell close to the experimental data (Fig. 2B). Also RMSD(Ω) improved from 15.2 to 0.94 Hz. The improvement of the correlation between the experimental and the calculated couplings after excluding the inconsistent data is shown in Fig. 3.

The plane of the peptide bond represents the most common example of a planar fragment in proteins. Residual dipolar couplings measured in the peptide

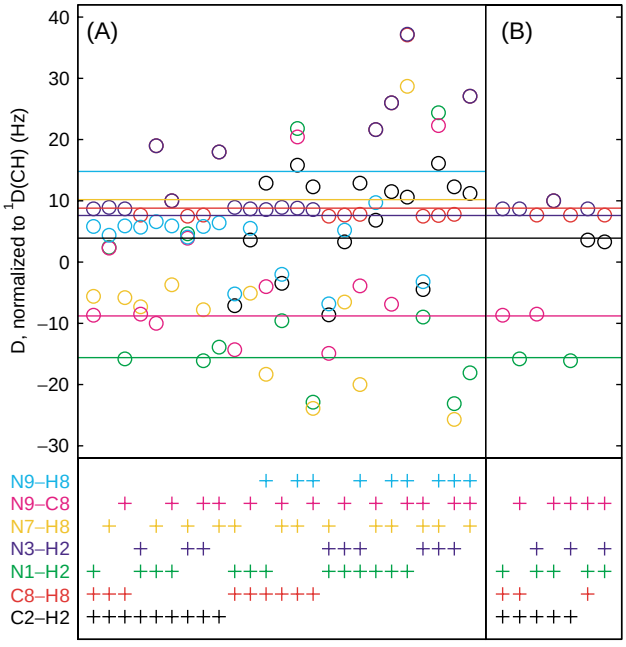

Fig. 2. Test of internal consistency of residual dipolar couplings in adenine 4 of the DNA hairpin displayed as described in Fig. 1. Panels A and B show the comparison of calculated and experimental values before and after excluding the inconsistent couplings, respectively. Symbols, lines, and labels of basis sets are color coded.

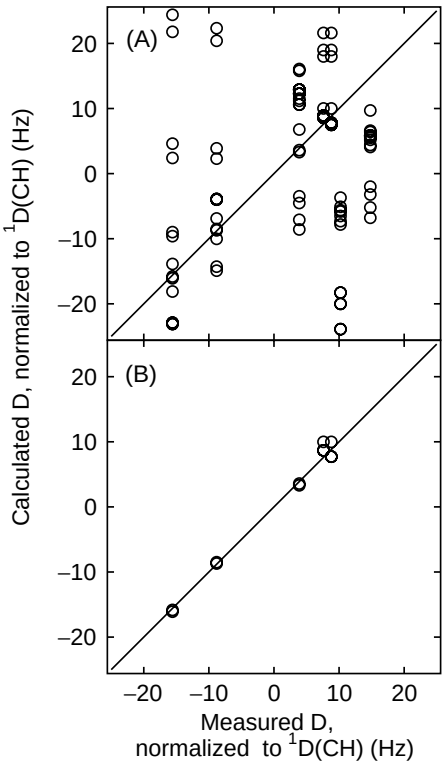

Fig. 3. Correlation of calculated and measured residual dipolar couplings in adenine 4 before (A) and after (B) excluding the inconsistent couplings.

bond plane of human ubiquitin were tested using the method proposed in this study. The experimental values were taken from the data published by Ottiger and Bax [12] for the charged phospholipid bicelles consisting of cetyltrimethylammonium bromide, dimyristoyl-phosphatidylcholine, and dihexanoyl-phosphatidylcholine in a molar ratio of 1:30:10. Internal consistency of four couplings measured in the peptide bond plane ($M = 4$) was tested in the same way as described for the DNA bases. Instead of showing the graphical output of the test for a single residue, correlation of the experimental data with the predicted values is presented in Fig. 4. The orientation of the relevant normalized internuclear vectors in the peptide plane is schematically indicated for each basis set in insets of Fig. 4. The plots contain data for all residues where the complete quartet of dipolar couplings was available. The correlation exhibits a strong dependence on the spatial distribution of the internuclear vectors corresponding to the dipolar couplings used as a basis set. It illustrates that attention should be paid to the selection of basis sets used for further data evaluation (the problem of the basis set selection is treated in Section 5). The data plotted for well-distributed basis set vectors (Panels A–C) show, that most experimental values provided a good correlation with the calculated expected values. Two values that exhibited the worst correlation are labeled in the

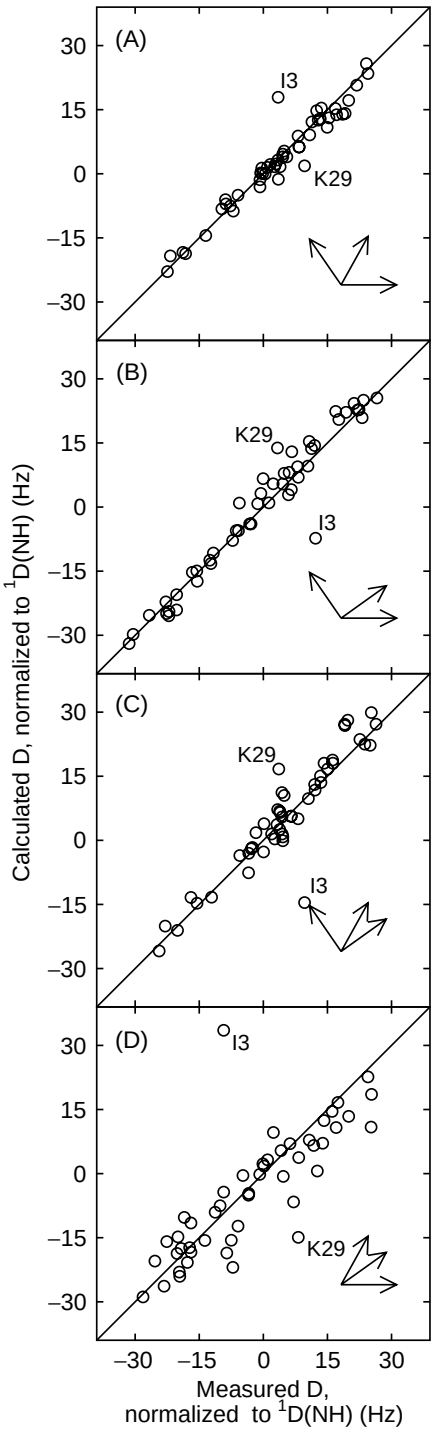

Fig. 4. Correlation of calculated and measured residual dipolar couplings ^{2}D(C′H^N) (A), ^{1}D(C′N) (B), ^{1}D(NH^N) (C), and ^{1}D(C$^\alpha$C′) (D) in peptide bond planes of ubiquitin. All couplings were normalized with respect to ^{1}D(NH^N). Relative orientation of the internuclear vectors corresponding to the used basis set is shown as an inset in each panel.

plot. The basis set used in Panel D exhibited worse correlation but the same inconsistent residues could be still easily identified. The effect of excluding several outliers present in the plots is more obvious when the experimental data are extended by observing additional second rank tensor interactions.

The use of second rank tensor interactions in structure determination is not limited to the most often used residual dipolar couplings. Cornilescu and Bax [4] reported values of chemical shift changes induced by partial alignment, $\Delta\delta$, for ^{1}H, ^{15}N, and carbonyl ^{13}C of ubiquitin peptide bonds. As the $\Delta\delta$ values were obtained in the same type of aligning media as the residual dipolar couplings discussed above, it was possible to test the internal consistency between the chemical shift and the dipolar coupling data. The small differences of the degree of alignment were accounted for by rescaling the measured couplings. The scaling factor was calculated as an average ratio of the ^{1}D(NH) values that were reported both by Ottiger and Bax [12] and by Cornilescu and Bax [4] (only the values of $|^1D(NH^N)|$ greater than 10 Hz were used). The expected $\Delta\delta$ and dipolar coupling values were then predicted for all acceptable combinations of experimental values using the peptide bond geometry and values of chemical shielding tensors reported by Cornilescu and Bax [4] (selection of the acceptable basis sets is described in Section 5). Note that the values of $\Delta\delta$ were also expressed as apparent coupling constants, obtained by multiplication of each $\Delta\delta_\lambda$ by the normalization factor of $\gamma B_0 c^D_{NH^N}/2\pi c^S_\lambda$.

The values of $\Delta\delta$ measured for three nuclei combined with the four sets of residual dipolar couplings provide a large array of data containing very detailed information about structure. At least four values per residue were available for 62 residues, providing the total number of 430 values. On the other hand, the correlation of the measured and calculated expected values, shown in Fig. 5A, indicated that not all data were mutually consistent. The correlation coefficient of 0.895 was obtained, nine residues exhibited RMSD(Ω) greater than 10 Hz, and additional 25 residues exhibited RMSD(Ω) in the range from 5 to 10 Hz. The internal consistency of the discussed data was improved in two steps. In the first step, 12 experimental values systematically deviating from the calculated predictions (average deviation larger than 10 Hz) were excluded. As a result, the number of calculated values decreased from 6948 to 6127 and the correlation coefficient increased from 0.895 to 0.939 (Fig. 5B). Although the general correlation improved, the data still contained inconsistent values (one residue with RMSD(Ω) greater than 10 Hz and 25 residues with RMSD(Ω) in the range from 5 to 10 Hz). In the second step, the limit of the average deviation of the experimental values was decreased from 10 to 5 Hz. Additional 38 experimental values were excluded in this step and one residue (Val 70) had to be excluded completely (only four experimental values were available for this residue). The selection resulted in a further decrease of the number of calculated values (to 4199) and in an increase of the correlation coefficient to 0.988 (Fig. 5C).

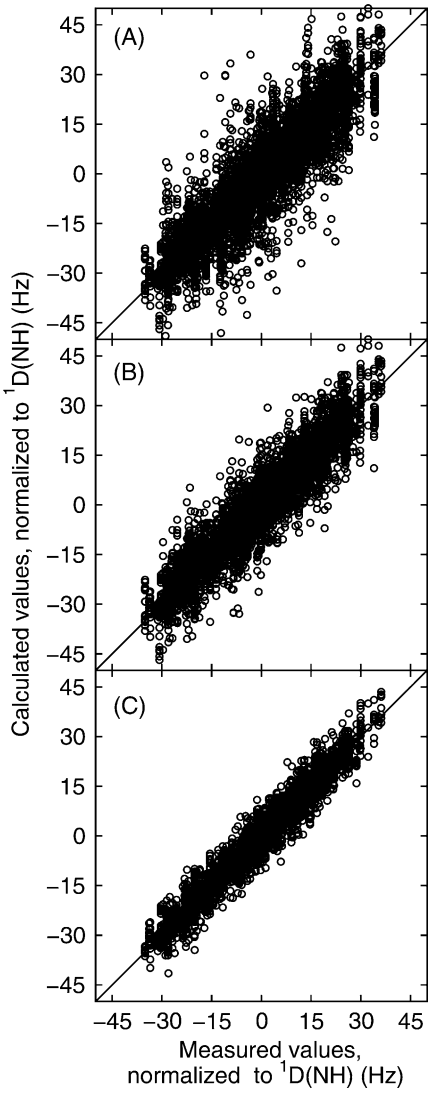

Fig. 5. Correlation of calculated and measured values of residual dipolar couplings and $\Delta\delta$ in peptide bond planes of ubiquitin, normalized with respect to ^{1}D(NH^N). Individual plots contain the complete set of experimental values (A), the experimental values without those deviating from the calculated values by more than 10 Hz (B), and the experimental values without those deviating from the calculated values by more than 5 Hz (C).

Only two residues exhibited RMSD(Ω) slightly exceeding 5 Hz (5.80 and 5.03 Hz). In most cases, the excluded values were chemical shift changes. Only two residual dipolar couplings exhibited deviation larger than 10 Hz (^{2}D(C′H^N) of Ile 3 and Lys 29, cf. Fig. 4). Several additional residual dipolar couplings deviated from their predicted values by more than 5 Hz, namely ^{1}D(NH^N) of Lys 29 and Leu 56, ^{1}D(C′N) of Phe 44 and Asn 60, ^{1}D(C$^\alpha$C′) of Leu 15, Lys 27, and Thr 55, and ^{2}D(C′H^N) of Glu 64.

The result of the internal consistency test shows that a significant improvement of the data quality can be obtained at the expense of a negligible reduction in the data amount. Although 12% of experimental values were discarded, the real loss of information is much lower because only three values are independent in planar fragments. With the exception of one excluded residue, the described procedure only represents discarding outliers from statistical sets.

5. Discussion

The internal consistency of measured dipolar couplings and chemical shift changes induced by partial alignment was tested on two types of planar molecular fragments (nucleic acid bases and peptide bond plane). Application of the method to the planar fragments requires that more than three experimental values are obtained for each fragment. In most cases, five or more values per fragment were available. It allowed to identify inconsistent items of data without loosing orientational information because discarding a single inconsistent value only corresponds to reduction of the data redundancy. Four experimental values were available for several fragments. In such cases, internal consistency of the data could be still tested (see Fig. 1). However, the complete set of four values had to be discarded when the test of the internal consistency failed, as it was in the case of Val 70 in ubiquitin. It demonstrates that the proposed method is useful even if the only data available are the dipolar couplings (^{1}D(NH^N), ^{1}D(C′N), ^{1}D(C$^\alpha$C′), and ^{2}D(C′H^N)). The demanding measurements of the chemical shift changes thus can be avoided at the price of loosing information about those residues for which inconsistency in the data was revealed.

Reliability of the proposed method depends on the errors in the calculated expected frequencies. There are two sources of random errors of the calculated values. The first source are the experimental errors in the measured frequencies. The second source are the errors in parameters used to determine the coefficient $b_{\mu\lambda}$, namely atomic coordinates, principal values of the shielding tensor, and its orientation with respect to the local geometry. When the method is applied to residual dipolar couplings measured in small fragments of well-defined geometry, the latter contribution is negligible compared to the experimental error in the measured couplings. The final error in a calculated expected frequency $\epsilon(\Omega_\mu^{\mathrm{calc}})$ is then mostly given by the experimental errors $\epsilon(\Omega_\lambda^{\mathrm{exp}})$ of L frequencies $\Omega_\lambda^{\mathrm{exp}}$ that are used to calculate the expected value $\Omega_{\mu\neq\lambda}^{\mathrm{calc}}$.

$$\epsilon(\Omega_\mu^{\mathrm{calc}}) = \pm\sqrt{\sum_{\lambda=1}^{L}(b_{\mu\lambda}\epsilon(\Omega_\lambda^{\mathrm{exp}}))^2}. \tag{16}$$

Two general conclusions can be drawn from Eq. (16). First, the method exhibits better precision for planar

fragments, where only three experimental errors ($L = 3$) contribute to the final error, than for fragments of a general shape with $L = 5$. Second, the error in the calculated frequency increases as the coefficients $b_{\mu\lambda}$ increase. The second conclusion is important for identifying and discarding basis sets that would lead to calculation of expected values with very high error.

The most exact selection criterion is based on the values of $\epsilon(\Omega_\mu^{\mathrm{calc}})$ defined by Eq. (16). It is also useful to define a general selection criterion that does not include the actual values of experimental errors. For such purpose, we introduce an experimental error propagation parameter E as

$$E = \sqrt{\sum_{\lambda=1}^{L} (b_{\mu\lambda})^2}. \tag{17}$$

The parameter E has a simple interpretation if one experimental error ϵ_{exp} can be used for all measured values. In such case, the error in the calculated values is equal to $E\epsilon_{\mathrm{exp}}$. In order to get a clearer insight into the error propagation, the definition of $b_{\mu\lambda}$ (Eq. (7)) can be expressed as a ratio of determinants $\det T^{(\mu\lambda)} / \det T$, where the matrix T is given by the tensor elements of interactions used as the basis and the matrix $T^{(\mu\lambda)}$ is obtained by replacing the λth row of T with the corresponding tensor elements of the μth interaction (Cramer's rule). Obviously, combinations of frequencies leading to small values of $\det T$ result in a large increase of the experimental errors. For example, it is the case if internuclear vectors corresponding to residual dipolar couplings measured in a planar fragment are collinear. For this reason, dipolar couplings measured between nuclei C8–H8 and N3–H2 in adenine could not be used in the same basis set as the corresponding internuclear vectors are almost collinear and $\det T$ approaches zero. It should be stressed that the intolerable increase of error is not limited to the extreme cases as discussed for the internuclear vectors C8–H8 and N3–H2. It is therefore advisable to limit the choice of basis sets to combinations of frequencies that do not produce errors of calculated experimental couplings too high compared to the range of the tested values. On the other hand, setting the E threshold too low should be avoided as well, as this would affect the reliability of the test by reducing the number of predicted values.

The effect of the error magnification is clearly visible in Fig. 4 where the correlation between the calculated and the experimental residual dipolar couplings reflects distribution of the internuclear vectors of individual basis sets in the peptide bond plane (the values of the parameter E in Panels A to D are 1.01, 1.63, 2.17, and 4.09, respectively). On the other hand, the same outliers can be identified in all four panels. The use of a basis set of $E = 4.09$ is therefore still acceptable.

In order to illustrate how to set the threshold for identifying the unacceptable basis sets, let us extend our discussion from four measured residual dipolar couplings to the full set of seven frequencies, tested in this study (four dipolar couplings and three chemical shift changes, $M = 7$). While only one basis set can be used to predict each value when only dipolar couplings are considered ($M = 4$), twenty basis sets can be used to predict each $\Omega_\mu^{\mathrm{calc}}$ when $M = 7$. Table 1 lists these basis sets for $\Omega_1^{\mathrm{calc}} = {}^1\mathrm{D}(\mathrm{C}^\alpha\mathrm{C}')$ of ubiquitin residue Met 1, in which all experimental values passed the test of internal consistency (after the unacceptable basis sets were excluded). Therefore, the calculated values of Ω_1^{calc}, also listed in Table 1, should not deviate too much from the normalized experimental value of ${}^1\mathrm{D}(\mathrm{C}^\alpha\mathrm{C}')$, −16.9 Hz. If we allow deviations $\Delta\Omega_1$ in a range of ±7.2 Hz, corresponding to approximately ±10% of the range of all normalized measured frequencies (−35.1 Hz to +36.2 Hz, see Fig. 5), only two values of Ω_1^{calc} exceed the limits. Similar results were obtained for the other six frequencies (data not shown). The lowest value of E leading to a deviation of $\Omega_\mu^{\mathrm{calc}}$ larger than 7.2 Hz from the experimental value ($\Delta\Omega_7 = -14.5$ Hz for $\Delta\delta(N)$) was found to be $E = 8.7$ (data not shown). The threshold for discarding unacceptable basis sets should be therefore lower than $E = 8.7$, but it should not be too low as it would exclude too many basis sets. Based on these considerations, we set the threshold to $E = 5$ when testing internal consistency of the ubiquitin data in this study. As a result, the basis sets had to be excluded only in 14 out of 140 calculations (when all seven experimental values were available). Note that a particular basis set may give $E < 5$ when used to calculate one predicted frequency but the same basis set may give $E > 5$ when used to calculate another frequency. It is thus not necessary to exclude such basis sets from calculating all four $(M - L)$ remaining frequencies.

The described selection of the acceptable basis sets, based on the parameter E, does not depend on the actual values of experimental errors. As the estimated errors were available for the data used in this study, $\epsilon(\Omega_\mu^{\mathrm{calc}})$ should provide another, and more exact, criterion for the basis set selection. However, when the $\epsilon(\Omega_\mu^{\mathrm{calc}})$ threshold was set so that the root mean square of the calculated and experimental error was equal to 7.2 Hz, corresponding to ±10% of the experimental range, three of the accepted basis sets resulted in larger than 20% deviations of $\Omega_\mu^{\mathrm{calc}}$ from the experimental values ($|\Delta\Omega_\mu| > 14.3$ Hz, see basis set 4 5 6 in Table 1). When the selection limit based on $\epsilon(\Omega_\mu^{\mathrm{calc}})$ was lowered to 5 Hz, corresponding to ±7% of the experimental range, 14 predicted values were discarded, 10 of them being identical to those excluded using $E = 5$ as a threshold. After excluding the unacceptable basis sets according to the lowered threshold, reasonable results of testing the internal consistency of the measured frequencies were

Table 1
Analysis of error propagation

Basis set									
λ_1	λ_2	λ_3	$b_{1\lambda_1}$	$b_{1\lambda_2}$	$b_{1\lambda_3}$	E	Ω_1^{calc}	$\epsilon(\Omega_1^{calc})$	$\Delta\Omega_1$
2	3	4	−3.0	2.2	1.8	4.1	−18.7	6.8	−1.8
2	3	5	−0.5	−0.4	−0.9	1.1	−18.0	1.0	−1.1
2	3	6	−1.5	0.8	−0.9	1.9	−20.2	3.5	−3.2
2	3	7	−1.6	1.6	1.2	2.5	−17.2	3.7	−0.3
2	4	5	−0.9	0.3	−0.8	1.2	−18.1	1.9	−1.2
2	4	6	−0.7	−1.0	−1.4	1.8	−21.0	1.8	−4.1
2	4	7	1.9	−4.5	4.3	**6.5**	-13.3	4.6	3.6
2	5	6	−0.8	−0.6	−0.3	1.1	−18.8	1.9	−1.8
2	5	7	−0.7	−0.7	0.3	1.0	−17.8	1.6	−0.9
2	6	7	−1.5	−1.7	−1.2	2.6	−23.1	3.5	−6.2
3	4	5	−0.9	−0.3	−1.1	1.5	−17.8	0.5	−0.9
3	4	6	−0.7	−1.9	−1.8	2.7	−21.7	1.2	−4.8
3	4	7	0.9	−2.0	2.6	3.4	−15.4	1.0	1.5
3	5	6	−0.9	−1.3	0.4	1.7	−17.1	0.6	−0.2
3	5	7	−1.2	−1.3	−0.5	1.9	−18.3	0.7	−1.4
3	6	7	−18.5	−21.9	−29.0	**40.8**	-92.0	**19.1**	−75.1
4	5	6	−7.6	4.0	−8.3	**12.0**	-35.7	5.1	−18.8
4	5	7	−1.2	−0.5	1.3	1.9	−16.6	0.5	0.3
4	6	7	−2.0	−1.0	1.2	2.5	−18.9	0.8	−2.0
5	6	7	−1.4	1.6	1.6	2.6	−13.0	1.1	3.9

Note. Propagation of error is shown for calculating $\Omega_1 = {}^1D(C^\alpha C')$ of ubiquitin Met 1 from basis sets constituted of three of the following normalized frequencies Ω_λ: $\lambda = 2$, ${}^2D(C'H^N)$; $\lambda = 3$, ${}^1D(C'N)$; $\lambda = 4$, ${}^1D(NH^N)$; $\lambda = 5$, $\Delta\delta(C')$; $\lambda = 6$, $\Delta\delta(H^N)$; and $\lambda = 7$, $\Delta\delta(N)$. The normalized experimental value of ${}^1D(C^\alpha C')$ was $\Omega_1^{exp} = (-16.9 \pm 0.9)$ Hz, while the experimental values of the remaining normalized frequencies were $\Omega_2^{exp} = (13.3 \pm 2.2)$ Hz, $\Omega_3^{exp} = (21.9 \pm 0.6)$ Hz, $\Omega_4^{exp} = (-15.5 \pm 0.2)$ Hz, $\Omega_5^{exp} = (2.6 \pm 0.1)$ Hz, $\Omega_6^{exp} = (19.7 \pm 0.6)$ Hz, and $\Omega_7^{exp} = (-25.6 \pm 0.3)$ Hz. $\Delta\Omega_1$ is defined as $\Omega_1^{calc} - \Omega_1^{exp}$. Values of E and $\epsilon(\Omega_1^{calc})$ that exceeded the limits of $E = 5$ and $\epsilon(\Omega_1^{calc}) = 7.1$ Hz are highlighted.

achieved. The same inconsistent values were excluded using the threshold of $\epsilon(\Omega_\mu^{calc}) = 5$ Hz or of $E = 5$ (data not shown), with one exception. The predicted value of $\epsilon(\Omega_7^{calc})$ for $\Delta\delta(N)$ calculated from $\Delta\delta(H^N)$, $\Delta\delta(C')$, and ${}^1D(NH^N)$ was still below the limit (3.8 Hz) but the values of $\Delta\delta(N)$ calculated using this basis set deviated by more than 10 Hz from their experimental values in almost all residues. The mentioned basis set had to be excluded when calculating Ω_7^{calc} in spite of its low $\epsilon(\Omega_7^{calc})$ in order to avoid false outliers. The problems of using $\epsilon(\Omega_\mu^{calc})$ probably reflect that the errors in the coefficients $b_{\mu\lambda}$ (Eq. (16)) were neglected and possibly that some experimental errors in the published data [4], especially those of the chemical shift changes, had been underestimated.

The comparison of two selection criteria, as presented above, leads to an important conclusion. The good performance of the method using the general criterion (parameter E) to select acceptable basis sets shows that the internal consistency of the measured values can be checked even when their experimental errors were not properly determined.

In addition to the discussed random errors, systematic errors can also affect the calculated expected frequencies. The systematic errors have been investigated for the most commonly used case of the residual dipolar couplings. The influence of the systematic experimental errors depends on the accuracy of methods used to measure the residual dipolar couplings. The coefficients $b_{\lambda\mu}$ may also introduce systematic errors if the used geometry does not correspond to the real shape of the fragment. Due to the discussed advantages of planar bodies, deviations from the planarity of such fragments are of special interest. In order to evaluate how a deviation of one internuclear vector from the plane of a fragment affects calculation of the corresponding coupling, it is useful to express Eq. (3) in the principal order frame of the tensor S, using spherical coordinates θ and ϕ [3,10]

$$\omega_\lambda = \frac{c_\lambda S_{z'z'}}{4}[(3\cos 2\theta + 1) - \eta(\cos 2\theta - 1)\cos 2\phi], \quad (18)$$

where the asymmetry parameter η is defined as $(S_{x'x'} - S_{y'y'})/S_{z'z'}$ and prime indicates principal values of the alignment tensor. It follows from Eq. (18) that the calculated values are most sensitive to angular changes of θ for $\theta = \pm\pi/4$. Assuming that a molecular fragment is oriented so that a calculated coupling ω_λ is most sensitive to deviations of the corresponding vector from the plane of the fragment, the change of the calculated coupling induced by rotation of the vector out of the plane can be expressed as

$$\Delta\omega_\lambda = \mp\frac{c_\lambda S_{z'z'}}{4}(3 - \eta\cos 2\phi)\sin 2\Delta\theta, \quad (19)$$

where $\Delta\theta$ is the angle between the internuclear vector and the plane of the fragment. Fig. 6 shows the impact

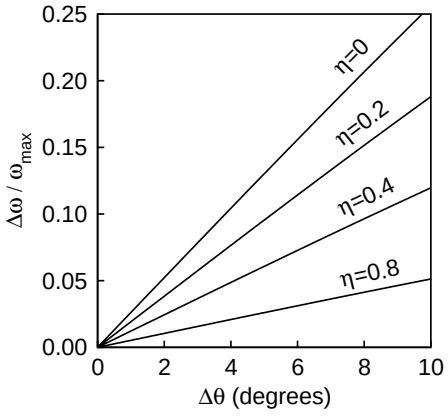

Fig. 6. Dependence of expected value of residual dipolar coupling calculated for a presumably planar fragments on deviation of the corresponding internuclear vector from the plane of the fragment.

of the out-of-plane deviations for several asymmetry parameters in the most sensitive orientation. The plot indicates that relatively moderate deviations from planarity may significantly affect the test if the molecular fragment is present in the most sensitive orientation. It is worthy to note that measurements of three-bond scalar couplings suggested that the root mean square deviation of the torsion angle ω are smaller than 5°, with the average value of $0 \pm 8°$ and $1 \pm 7°$ in α-helical and β-sheet residues, respectively [6]. Such deviations are likely to introduce observable errors but not to deteriorate the proposed test of the internal consistency.

6. Conclusions

The described method allows to test measured values of the second rank tensor interactions in partially aligned biomacromolecules. If the number of frequencies measured in a rigid fragment of the molecule is larger than the number of linearly dependent frequencies (three for a planar fragment), it is possible to assess whether the measured values are consistent with one another (and with the geometry of the fragment). The information about the internal consistency is useful when deciding which experimental data can be used in reliable structure calculations. The test requires only sufficient number of experimental data and a knowledge of local spatial dependence of the observed interactions (such as local geometry in the case of residual dipolar couplings or principal values and orientation of shielding tensor in the case of anisotropic contributions to the chemical shift). No information about alignment of the studied molecule is required. The method is best suited for planar fragments and has been successfully tested on residual dipolar couplings and anisotropic chemical shifts obtained for protein peptide bonds and nucleic acid bases.

Acknowledgments

This work was supported by Grant No. LN00A016 from Ministry of Education of the Czech Republic. The authors also thank M.L. Munzarová, D. Munzar, and R. Fiala for helpful discussions.

References

[1] E. Brunner, Residual dipolar couplings in protein NMR, Concepts Magn. Reson. 13 (2001) 238–259.
[2] W.Y. Choy, M. Tollinger, G.A. Mueller, L.E. Kay, 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 (2001) 31–40.
[3] 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.
[4] G. Cornilescu, A. Bax, Measurement of proton, nitrogen, and carbonyl chemical shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase, J. Am. Chem. Soc. 122 (2000) 10143–10154.
[5] E. de Alba, N. Tjandra, NMR dipolar couplings for the structure determination of biopolymers in solution, J. Biomol. NMR 40 (2002) 175–197.
[6] J. Hu, A. Bax, Determination of ϕ and χ_1 angles in proteins from ^{13}C–^{13}C three-bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond?, J. Am. Chem. Soc. 119 (1996) 6360–6368.
[7] J.C. Hus, D. Marion, M. Blackledge, Determination of protein backbone structure using only residual dipolar couplings, J. Am. Chem. Soc. 123 (2001) 1541–1542.
[8] J.A. Losonczi, M. Andrec, M.W.F. Fischer, J.H. Prestegard, Order matrix analysis of residual dipolar couplings using singular value decomposition, J. Magn. Reson. 138 (1999) 334–342.
[9] E.T. Mollova, A. Pardi, NMR solution structure determination of RNAs, Curr. Opin. Struct. Biol. 10 (2000) 298–302.
[10] S. Moltke, S. Grzesiek, Structural constraints from residual tensorial couplings in high resolution NMR without an explicit term for the alignment tensor, J. Biomol. NMR 15 (1999) 77–82.
[11] G.A. Mueller, W.Y. Choy, D.W. Yang, J.D. Forman-Kay, R.A. Venters, L.E. Kay, Global folds of proteins with low densities of NOE using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein, J. Mol. Biol. 300 (2000) 197–212.
[12] M. Ottiger, A. Bax, Determination of relative N–H^N, N–C′, C^α–C′, and C^α–H^α effective bond lengths in a protein by NMR in a dilute liquid crystalline phase, J. Am. Chem. Soc. 120 (1998) 12334–12341.
[13] P. Padrta, R. Štefl, L. Králík, L. Žídek, V. Sklenář, Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings, J. Biomol. NMR 24 (2002) 1–14.
[14] P. Permi, A. Annila, Transverse relaxation optimised spin-state selective NMR experiments for measurement of residual dipolar couplings, J. Biomol. NMR 16 (2000) 221–227.
[15] B. Simon, M. Sattler, De novo structure determination from residual dipolar couplings by NMR spectroscopy, Angew. Chem.-Int. Ed. 41 (2002) 437–440.

[16] J.R. Tolman, Dipolar couplings as a probe of molecular dynamics and structure in solution, Curr. Opin. Struct. Biol. 11 (2001) 532–539.
[17] L. Žídek, R. Štefl, V. Sklenář, NMR methodology for the study of nucleic acids, Curr. Opin. Struct. Biol. 11 (2001) 275–281.
[18] L. Žídek, H.H. Wu, J. Feigon, V. Sklenář, Measurement of small scalar and dipolar couplings in purine and pyrimidine bases, J. Biomol. NMR 21 (2001) 153–160.
[19] Z. Wu, N. Tjandra, A. Bax, ^{31}P chemical shift anisotropy as an aid in determining nucleic acid structure in liquid crystals, J. Am. Chem. Soc. 123 (2001) 3617–3618.
[20] J.L. Yan, T. Corpora, P. Pradhan, J.H. Bushweller, MQ-HCN-based pulse sequences for the measurement of ^{13}C1′–^{15}N, ^{1}H1′–^{15}N, ^{13}C1′–^{13}C2′, ^{1}H1′–^{13}C2′, ^{13}C6/8–^{15}N6/8, ^{1}H6/8–^{15}N, ^{13}C6–^{13}C5, ^{1}H6–^{13}C5 dipolar couplings in ^{13}C, ^{15}N-labeled DNA (and RNA), J. Biomol. NMR 22 (2002) 9–20.
[21] D.W. Yang, R.A. Venters, G.A. Mueller, W.Y. Choy, L.E. Kay, TROSY-based HNCO pulse sequences for the measurement of ^{1}HN–^{15}N, ^{15}N–^{13}CO, ^{1}HN–^{13}CO, ^{13}CO–^{13}C$^{\alpha}$ and ^{1}HN–^{13}C$^{\alpha}$ dipolar couplings in ^{15}N, ^{13}C, ^{2}H-labeled proteins, J. Biomol. NMR 14 (1999) 333–343.