J Biomol NMR (2008) 40:71–81
DOI 10.1007/s10858-007-9210-6

ARTICLE

Influence of the fluctuations of the alignment tensor on the analysis of the structure and dynamics of proteins using residual dipolar couplings

X. Salvatella · B. Richter · M. Vendruscolo

Received: 20 September 2007 / Accepted: 1 November 2007 / Published online: 21 November 2007

Abstract It has been suggested that the fluctuations of the alignment tensor can affect the results of procedures for characterizing the structure and the dynamics of proteins using residual dipolar couplings. We show here that the very significant fluctuations of the steric alignment tensor caused by the dynamics of proteins can be safely ignored when they do not correlate with those of the bond vectors. A detailed analysis of these correlations in the protein ubiquitin reveals that their effects are negligible for the analysis of backbone motions within secondary structure elements, but also that they may be significant in turns, loops and side chains, especially for bond vectors that have small residual dipolar couplings. Our results suggest that methods that explicitly consider the motions of the alignment tensor will be needed to study the large-scale structural fluctuations that take place on the millisecond timescale, which are often important for the biological function of proteins, from residual dipolar coupling measurements.

Keywords Protein dynamics · Alignment tensor · Residual dipolar couplings · Correlated motions

Introduction

Since the first measurement of residual dipolar couplings (*RDCs*) in proteins using an external alignment medium (Tjandra and Bax 1997), procedures to refine the average structure of biological macromolecules using this NMR observable have become standard tools in structural biology (de Alba and Tjandra 2002; Bax 2003). In addition, the ability of *RDCs* to report on protein dynamics was promptly recognized, attracting great interest because *RDCs* are sensitive to motions up to the millisecond timescale, which are important for biological processes such as enzymatic catalysis and allosteric communication (Tolman et al. 1997; Tolman et al. 2001; Meiler et al. 2001; Blackledge 2005; Bouvignies et al. 2006; Bax and Grishaev 2005). Various approaches have been proposed to extract the amplitude of bond vector motions on this timescale from *RDC* measurements. *RDCs* measured in multiple alignment media have been used to describe the motions of the bond vectors of the protein in a model-free manner (Meiler et al. 2001; Peti et al. 2002) and within models of various degrees of complexity (Meiler et al. 2001; Bernadó and Blackledge 2004; Bouvignies et al. 2005). A complementary method makes use of restrained molecular simulations to generate ensembles that represent the range of conformations adopted by the protein with no a priori assumption about the type of motions present other than the use of a molecular mechanics force-field (Clore and Schwieters 2004a, b).

Since *RDCs* have the potential to provide high-resolution information about the dynamics of proteins and their associated biological function, it is important to establish the range of validity of the different approaches that can be used to analyze them. In this work, we focus on the analysis of model-free methods that extract the amplitude of bond vector fluctuations, which have been used to study the dynamics of ubiquitin (Meiler et al. 2003; Lakomek et al. 2006) and protein G (Bernadó and Blackledge 2004; Bouvignies et al. 2005). Since the alignment tensor is

X. Salvatella (✉) · B. Richter · M. Vendruscolo (✉)
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
e-mail: xs210@cam.ac.uk

M. Vendruscolo
e-mail: mv245@cam.ac.uk

assumed to be invariant in these approaches, we analyze under which circumstances the fluctuations that it experiences (Louhivuori et al. 2006) affect the analysis of the structure and the dynamics of proteins. We carry out this analysis by using molecular dynamics (MD) simulations to generate a native state ensemble for the protein ubiquitin, which is similar to experiment in terms of the observed motions on the sub-nanosecond timescale, and by calculating the steric alignment tensor of the members of the ensemble with the PALES program (Zweckstetter and Bax 2000; Zweckstetter et al. 2004). Although we find that the alignment tensor experiences very significant fluctuations, we show that these can only affect the estimation of protein dynamics from *RDCs* if they are correlated to those of the relevant bond vectors. We then analyze these correlations for ubiquitin and show, as previously suggested (Tolman and Ruan 2006), that they are significant only for the most flexible regions of the protein backbone such as the loops, the termini and the side chains.

Results

Impact of protein dynamics on the interpretation of RDCs

Since experimental *RDCs* are averaged across the range of conformations that a protein can adopt, NMR measurements yield ensemble-averaged values of the *RDCs* (Blackledge 2005)

$$\begin{aligned} D^{\exp} = \langle D \rangle &= \left\langle -\frac{\mu_0 \gamma_P \gamma_Q h}{8\pi^3 r^3} \sum_{ij} S_{ij} \cos\phi_i \cos\phi_j \right\rangle \\ &\approx -\frac{\mu_0 \gamma_P \gamma_Q h}{8\pi^3 \langle r^3 \rangle} \sum_{ij} \langle S_{ij} \cos\phi_i \cos\phi_j \rangle \\ &= C_{PQ} \sum_{ij} \langle S_{ij} \cos\phi_i \cos\phi_j \rangle \end{aligned} \quad (1)$$

where $i,j = (x,y,z)$, ϕ_i is the angle between the interatomic vector PQ and the axis i of the fixed molecular frame, S_{ij} is the ij element of the alignment tensor $\boldsymbol{S}$, which is a real symmetric traceless 3×3 matrix that expresses the degree and main directions of alignment, γ_X is the gyromagnetic ratio of nucleus X, h is Planck's constant, μ_0 is the magnetic susceptibility of vacuum and r is the length of the interatomic vector PQ.

The angular brackets indicate an ensemble average over the different conformations of the molecules in solution, and Eq. 1 exploits the fact that the fluctuations of the lengths and directions of the interatomic vectors occur on different timescales and are thus not significantly correlated. An additional and important assumption made in this type of analysis is that the structure and dynamics of the proteins not interacting with the alignment medium are similar to those of the proteins in its proximity, which are the ones that are actually probed experimentally.

By assuming that the alignment tensor of the protein remains constant on the timescale sampled by *RDCs*, it is possible to rewrite Eq. 1 in the eigenframe of the tensor using the spherical coordinates r, θ and φ (Tjandra and Bax 1997)

$$\begin{aligned} D^{\exp} &\approx C_{PQ} \sum_{ij} S_{ij} \langle \cos\phi_i \cos\phi_j \rangle \\ &= C_{PQ} \frac{S'_{zz}}{2} \left[\langle 3\cos^2\theta - 1 \rangle + \frac{3R}{2} \langle \sin^2\theta \cos 2\varphi \rangle \right] \end{aligned} \quad (2)$$

where S'_{xx}, S'_{yy} and S'_{zz} are the eigenvalues of S, the axes x, y and z are chosen so that $|S'_{zz}| \geq |S'_{yy}| \geq |S'_{xx}|$ and $R = (2/3)\cdot(S'_{xx} - S'_{yy})/S'_{zz}$.

As shown by Meiler et al. (2001) the functional dependence of $D^{\exp}$ on the polar angles of each bond vector can be recast in terms of spherical harmonics $Y_{2M}(\theta,\varphi)$. The measurement of the *RDCs* in at least 5 independent alignment media allows the accurate characterization of the spherical harmonics in the molecular frame, provided that good estimates of S'_{zz} and R can be obtained from the experimental data. Since the spherical harmonics fully describe the angular motions of the bond vectors, they can be used to compute a model-free order parameter, S^2_{RDC}, which is mathematically equivalent to the order parameter obtained from the analysis of heteronuclear relaxation rates S^2_{LS} (Lipari and Szabo 1982) but which reports on the motions sampled over a longer timescale

$$S^2_{RDC} = \frac{4\pi}{5} \sum_{M=-2}^{2} \langle Y_{2M}(\theta,\varphi) \rangle \langle Y^*_{2M}(\theta,\varphi) \rangle \quad (3)$$

In addition to this model-free characterization of protein dynamics, the spherical harmonics can be fit to motional models of different degrees of complexity to provide estimates for the angular amplitudes of the fluctuations of bond vectors (Bernadó and Blackledge 2004; Bouvignies et al. 2005) and improved average structures.

Analysis of the fluctuations of the alignment tensor using MD simulations

To assess the consequences of the assumption that the alignment tensor of the protein is invariant, we studied its fluctuations using MD simulations of ubiquitin (Richter et al. 2007). We generated a 22 ns trajectory at 300 K using the CHARMM program with the all-atom CHARMM22 forcefield (Brooks et al. 1983) in explicit solvent starting from the minimized X-ray structure of the protein (Vijay-Kumar et al.

1987). After discarding the first 2 ns for equilibration, we superimposed the resulting 2000 structures by minimizing their root mean square deviation (*rmsd*) to the average structure and computed the steric alignment tensor for each protein conformation using PALES (Zweckstetter and Bax 2000). PALES calculates the alignment tensor by sampling a large number of protein orientations and rejecting those where the atoms of the protein, which is considered rigid, clash with an infinitely long wall that models the steric alignment medium; this and similar computer programs have been shown to be remarkably accurate at predicting the *RDCs* of globular (Zweckstetter and Bax 2000) and unfolded proteins (Jha et al. 2005; Bernadó et al. 2005) in steric alignment. We, as others (Louhivuori et al. 2006), have found that the elements of the tensor do indeed fluctuate very significantly in the native ensemble; the results of this analysis are shown in Fig. 1, where it is apparent that all tensor elements have unimodal distributions with similar widths.

As previously pointed out (Tolman and Ruan 2006; Lakomek et al. 2006), the presence of these fluctuations does not necessarily imply that it is not possible to accurately derive protein dynamics from *RDCs*. By neglecting the correlations between the motions of the alignment tensor elements S_{ij} and those of the bond vector, which enter into the *RDC* calculations as $\cos\phi_i\cos\phi_j$ (Eq. 1), it is possible to provide an approximate expression, denoted by a*, for $D^{\exp}$

$$D^{\exp,*} = C_{PQ}\sum_{ij}\langle S_{ij}\rangle\langle\cos\phi_i\cos\phi_j\rangle = C_{PQ}\frac{S^*_{zz}}{2}\left[\langle 3\cos^2\theta - 1\rangle + \frac{3R^*}{2}\langle\sin^2\theta\cos 2\varphi\rangle\right] \quad (4)$$

where, S^*_{xx}, S^*_{yy} and S^*_{zz} are the eigenvalues of the ensemble-averaged alignment tensor $<\boldsymbol{S}>$, and $R^* = (2/3)\cdot(S^*_{xx} - S^*_{yy})/S^*_{zz}$.

If the correlations are truly absent, then $D^{\exp} = D^{\exp,*}$, and the motions of the tensor do not affect the determination of the average structure and the amplitudes of the fluctuations of the bond vectors from *RDCs*. It is important, however, to emphasize that in order to apply Eq. 4 it is necessary to very accurately determine the average alignment tensor $<\boldsymbol{S}>$; to achieve this result, the experimental *RDCs*, $D^{\exp}$, are fit to the average structure, previously obtained independently, using a singular value decomposition (SVD) procedure (Zweckstetter and Bax 2000). This procedure identifies $<\boldsymbol{S}>$ as the alignment tensor S^{fit} that minimizes the quality factor Q

$$Q = \frac{rms(D^{av}_{SVD} - D^{\exp})}{rms(D^{\exp})} \quad (5)$$

with

$$D^{av} \approx C_{PQ}\sum_{ij} S^{fit}_{ij}\cos\phi^{av}_i\cos\phi^{av}_j \quad (6)$$

When the assumption about the absence of correlations holds, this procedure is very accurate if the number of bond vectors used for the fitting is large compared to 5, the number of independent elements of the alignment tensor. In these cases, although the quality of the estimation of $<S_{ij}>$ from S^{fit}_{ij} depends on the model of the average structure (Zweckstetter and Bax 2002), its choice has an almost negligible effect on the outcome of the analysis of bond vector dynamics (Lakomek et al. 2006).

Thus, in summary, the main assumptions of model-free methods for the extraction of protein dynamics from *RDCs* are (i) that the motions of the tensor and the bond vectors are uncorrelated and (ii) that it is possible to accurately compute the ensemble-averaged alignment tensor of the protein from the knowledge of the experimental *RDCs* and of the average structure.

Analysis of the correlations between S_{ij} and $cos\phi_i cos\phi_j$

Since the assumption about the absence of correlation between the motions of the alignment tensor and those of the bond vectors is explicit in the formulation of both the spherical harmonics and the equations that describe the motional models, it is crucial to analyze the presence of correlations in the study of the structure and dynamics of the backbone and side chains using *RDCs*. In this section, we describe in detail the dynamics of eight representative bond vectors of ubiquitin (the *NH* bond vectors of residues

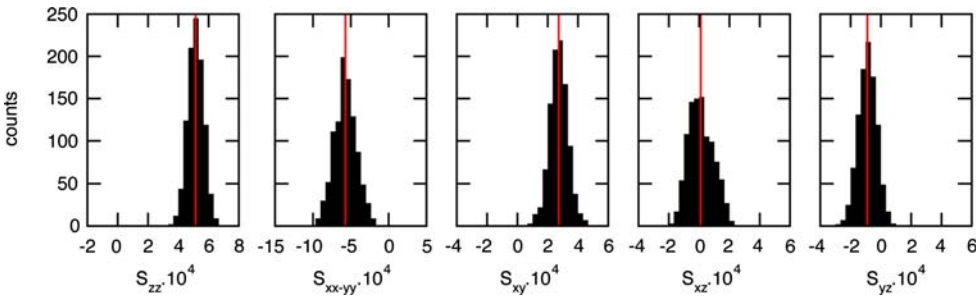

Fig. 1 Fluctuations on the sub-nanosecond timescale of the five independent elements of the alignment tensor in the fixed molecular frame in a 20 ns MD trajectory of the protein ubiquitin. The vertical red lines indicate the average values of the elements of the alignment tensor and the black histogram the distributions of their values

Val5, Thr9, Ile30 and Arg74 and the $C_\beta C_\gamma$ bond vectors of Val5, Lys48, Leu56 and Leu71), which have been selected because they belong to different elements of secondary structure (β-strand 1 for Val5, the turn connecting β strands 1 and 2 for Thr9, the α-helix for Ile30, the long loop between β-strand 3 and the first 3_{10}-helix for Lys48, the second 3_{10}-helix for Leu56, and the disordered C-terminus for Leu71 and Arg74) and fluctuate with significantly different amplitudes. The fluctuations of these bond vectors in the six *ij* combinations defined in a fixed reference frame are shown in Figs. 2 and 3, where we present histograms of the product of $\cos\phi_i\cos\phi_j$ since this is the quantity that relates the direction of bond vector *PQ* to the *RDC* in the fixed molecular frame (see Eq. 1). Inspection of the distributions of $\cos\phi_i\cos\phi_j$ shows that their widths can vary significantly for a given bond vector depending on the *ij* combination considered. For example, the distribution of $\cos^2\phi_x^{NH}$ for Val5 is narrow whereas those of $\cos^2\phi_y^{NH}$ and $\cos^2\phi_z^{NH}$ for the same residue are very wide. As expected, there is a marked difference between the distributions of residues in regular secondary structure elements, such as Val5 and Ile30, for which the distributions are relatively narrow for most *ij* combinations, and for residues in the most flexible region of the protein, such as Thr9 and Arg74, which exhibit wide distributions in many cases. For the $C_\beta C_\gamma$ bond vectors, we observe that residues Lys48 and Leu71, which are exposed to the solvent, have distributions of $\cos\phi_i\cos\phi_j$ that are multimodal because these side chains present multiple rotameric states in the ensemble; this behavior contrasts that observed for Val5 and Leu56, which are buried in the structure of the protein and present unimodal distributions with a width comparable to that of the *NH* bond vectors of secondary structure elements.

We then studied the degree of correlation ρ_{ij} between S_{ij} and $\cos\phi_i\cos\phi_j$ in the fixed reference frame for all residues of ubiquitin, with a special emphasis on the eight bond vectors presented in the previous paragraph. An absence of correlation would imply that it is correct to ignore the fluctuations of S_{ij} presented in Fig. 1; however, the presence of significant correlations would affect the results of any approach that ignores the fluctuations of S_{ij} for the analysis of protein structure and dynamics using *RDCs*. We computed the Pearson linear correlation coefficient, ρ_{ij}, between the distributions of $\cos\phi_i\cos\phi_j$ and S_{ij} in the ensemble for each *NH* and $C_\beta C_\gamma$ bond vector of ubiquitin and obtained the results presented in Figs. 4 and 5, where the residues discussed in the previous paragraph are indicated as red dots. An analysis of the results obtained for the *NH* bond vectors indicates that, for most residues, the correlation coefficient is lower than 0.2. However, there are two stretches of the polypeptide chain for which the motions of the bond vectors are significantly correlated to those of the elements of the alignment tensor—the turn between β-strands 1 and 2 (residues Leu8 to Gly10) and the C-terminus of the protein (residues Leu71 to Gly76), where higher values of $|\rho|$ are obtained, particularly for element *xz*. The degree of correlation of the $C_\beta C_\gamma$ bond vectors is found to be higher than that for the *NH* bond vectors for all *ij* combinations (Figs. 4 and 5). However, contrary to what we found for *NH* bond vectors, we observe that residues close in sequence do not tend to have similar values of ρ, suggesting that the degree of correlation for a given residue

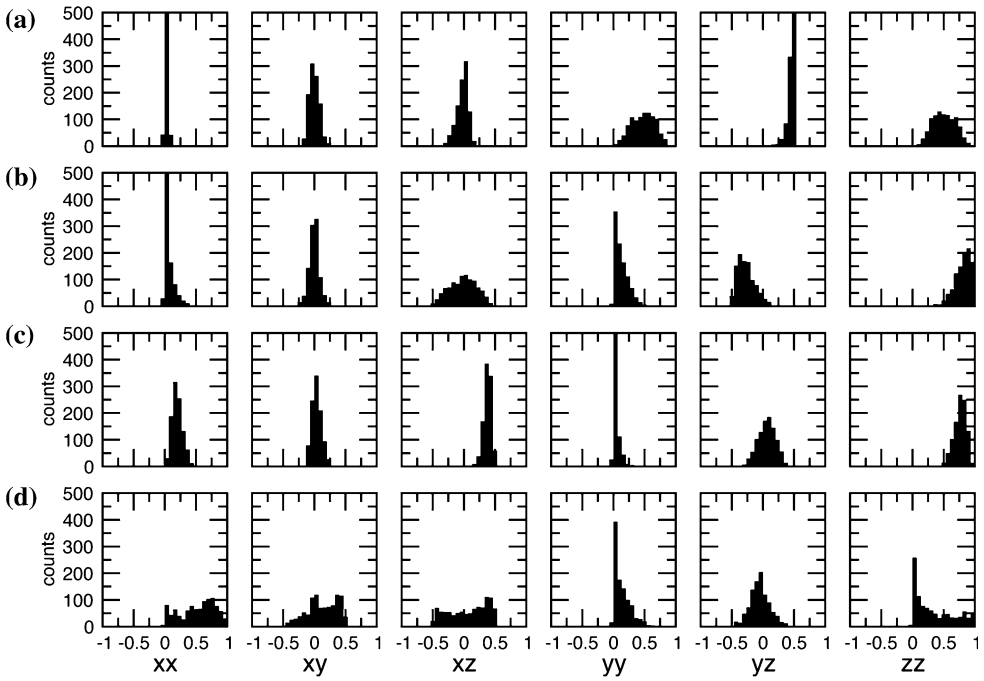

Fig. 2 Fluctuations of selected *NH* bond vectors in the sub-nanosecond timescale. Distribution of values of $\cos\phi_i\cos\phi_j$ in the fixed molecular frame for four different residues of ubiquitin: (**a**) Val5 (**b**) Thr9 (**c**) Ile30 (**d**) Arg74

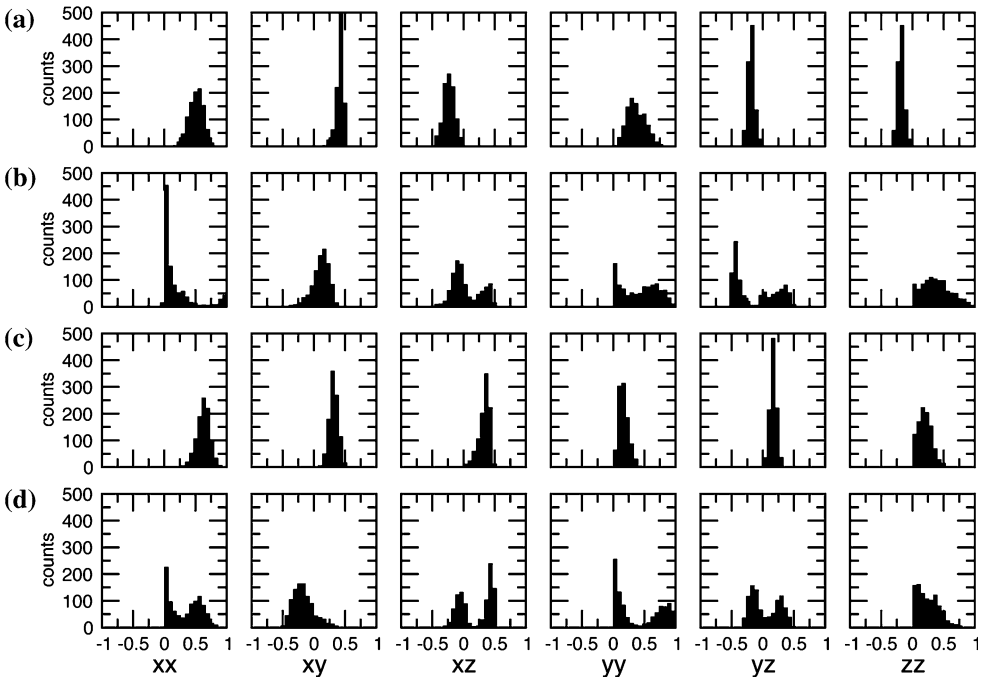

Fig. 3 Fluctuations of selected $C_\beta C_\gamma$ bond vectors in the sub-nanosecond timescale. Distribution of values of $\cos\phi_i\cos\phi_j$ in the fixed molecular frame for four different residues of ubiquitin: (**a**) Val5 (**b**) Lys48 (**c**) Leu56 (**d**) Leu71

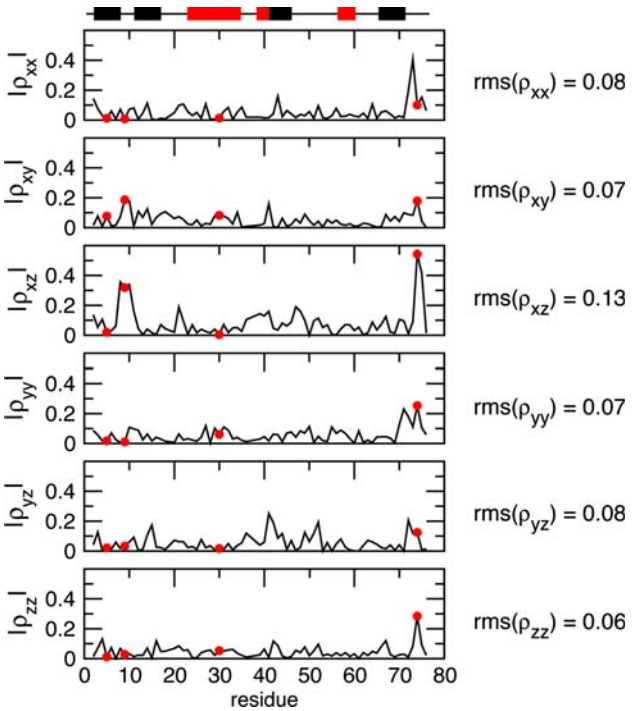

Fig. 4 Correlation coefficient $|\rho_{ij}|$ between S_{ij} and $\cos\phi_i\cos\phi_j$ for the 6 different directions of the alignment tensor; the (rms) root mean square of ρ_{ij} is shown next to the plot. The residues shown as red dots are those for which the fluctuations of the *NH* bond vector are also shown in Fig. 2; in these plots the elements of secondary structure are shown as colored rectangles (β-strands are shown in black and α-helices are shown in red)

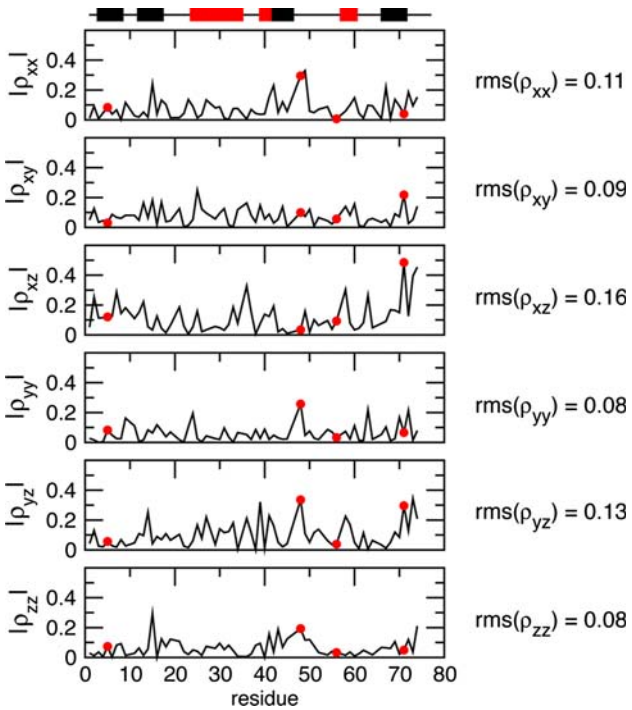

Fig. 5 Correlation coefficient $|\rho_{ij}|$ between S_{ij} and $\cos\phi_i\cos\phi_j$ for the 6 different directions of the alignment tensor; the (rms) root mean square of ρ_{ij} is shown next to the plot. The residues shown as red dots are those for which the fluctuations of the $C_\beta C_\gamma$ bond vector are also shown in Fig. 3; in these plots the elements of secondary structure are shown as colored rectangles (β-strands are shown in black and α-helices are shown in red)

is at least in part determined by the actual properties of the side chain rather than by the secondary structure or flexibility of the residue.

In Fig. 6, we present correlation plots for the four different *NH* bond vectors that we have chosen as representative in the fixed reference frame. For relatively rigid residues such as Val5, found in β-strand 1 of ubiquitin, we found essentially no correlation; this is particularly clear for elements *xx* ($\rho = -0.01$) and *xz* ($\rho = 0.02$); a very similar situation is found for Ile30 of the α-helix of ubiquitin. The presence of correlations is, however, appreciable for Thr9 ($\rho_{xy} = 0.19$, $\rho_{xz} = 0.32$) and, especially, for Arg74 ($\rho_{xz} = 0.54$, $\rho_{zz} = 0.28$), where it is already clear by visual inspection of the plots presented in Fig. 6d that the motions of the alignment tensor and those of the bond vector are significantly correlated. Therefore the assumption of absence of correlations appears to be correct for the *NH* bond vectors of residues Val5 and Ile30, which are found in well-defined elements of secondary structure, but incorrect for residues in flexible regions of the protein molecule such as turns (Thr9) and especially unstructured termini (Arg74). For the $C_\beta C_\gamma$ bond vectors, we find that the motions of Val5 and Leu56, which are relatively rigid and do not contribute to the surface of the protein, are weakly correlated to those of the elements of the alignment tensor as shown in Fig. 7. By contrast, solvent-exposed and very dynamic residues tend to present a significant degree of correlation (Lys48, $\rho_{xx} = -0.30$ and $\rho_{yz} = -0.34$; Leu71, $\rho_{xz} = -0.49$ and $\rho_{yz} = -0.30$), because the different rotameric states available to the side chain directly affect the value of S_{ij}.

Effects of the correlations on the determination of the average structure of proteins from RDCs

In this section, we discuss how the correlations presented above can in principle affect the results of the *RDC*-based determination of the average structure of proteins because they affect the relationship between the experimental observable, $D^{\exp}$, and the underlying distribution of structures that contribute to it.

Since standard structure determination protocols (Schwieters et al. 2003) assume that the experimental *RDC* can be expressed using one single alignment tensor, any difference $\Delta D^{\exp}$, between the *RDC* computed by assuming absence of correlations, $D^{\exp,*}$, and its exact counterpart, $D^{\exp}$

$$\Delta D^{\exp} = D^{\exp} - D^{\exp,*} \tag{7}$$

may introduce an error in the average orientation of the bond vector. We investigate the size of the error $|\Delta D^{\exp}|$ in the cases of the *NH* and $C_\beta C_\gamma$ bond vectors of ubiquitin (Fig. 8). Since not all *ij* combinations carry the same weight in the sum giving the experimental *RDC* (Eq. 1), it is not straightforward to uncover a relationship between the

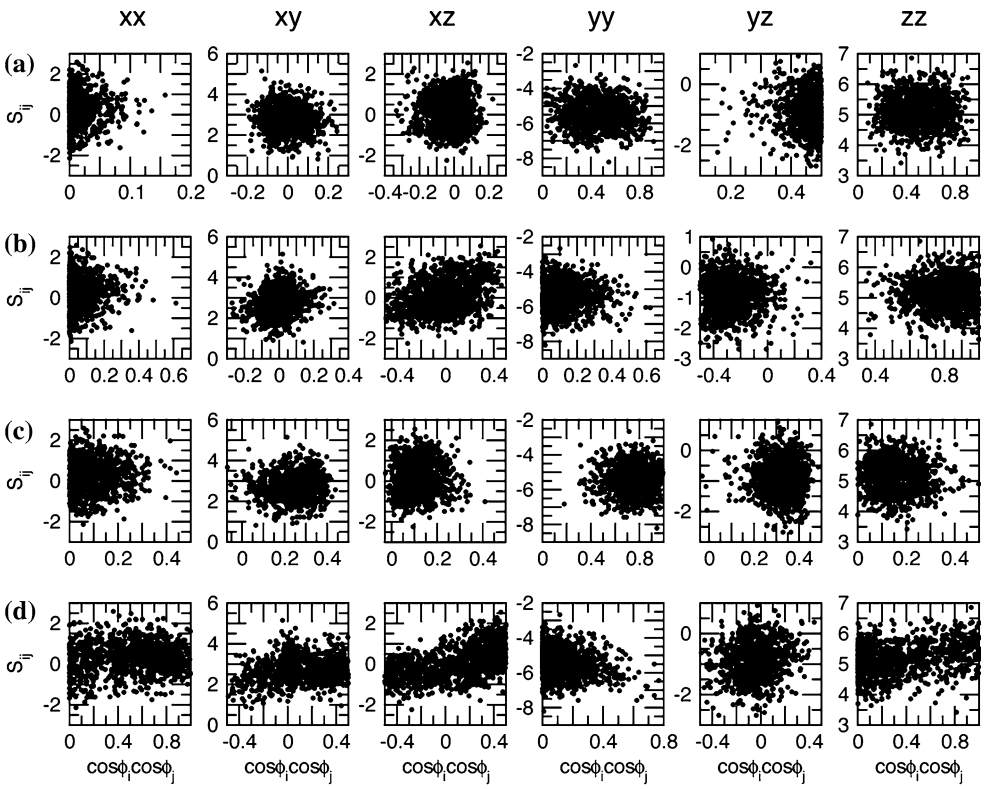

Fig. 6 Correlation between the fluctuations of the elements of the alignment tensor S_{ij} and those of four representative backbone bond vectors expressed as $\cos\phi_i^{NH}\cos\phi_j^{NH}$ in the sub-nanosecond timescale for four different residues of ubiquitin: (**a**) Val5 (**b**) Thr9 (**c**) Ile30 (**d**) Arg74

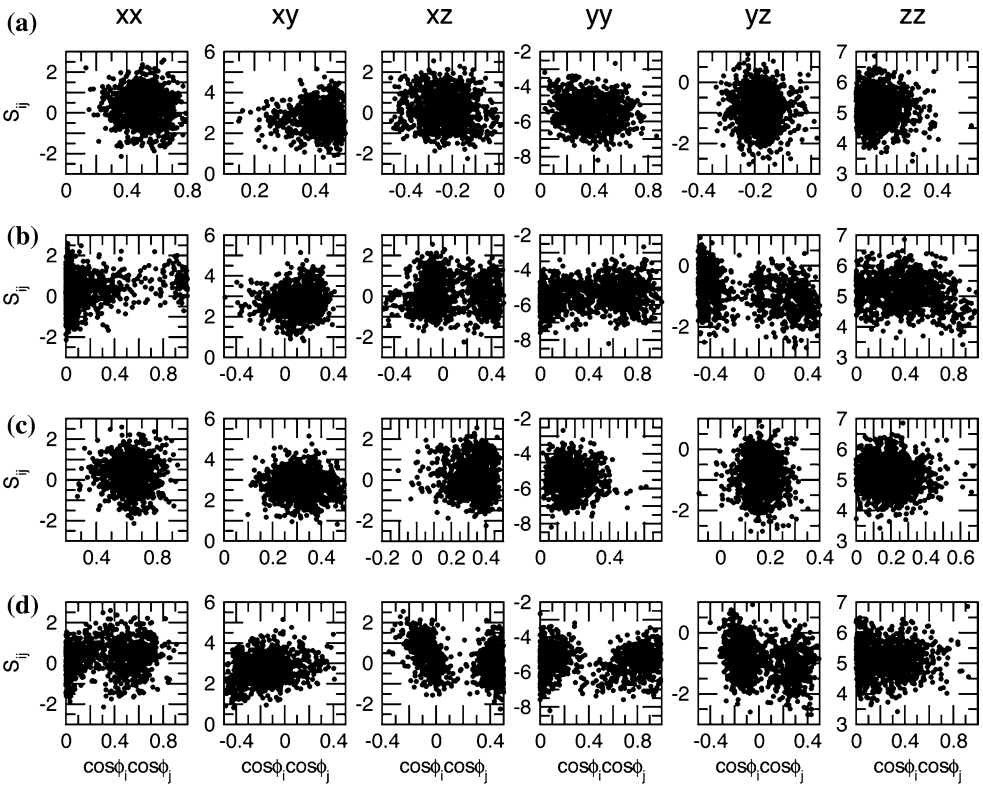

Fig. 7 Correlation between the fluctuations of the elements of the alignment tensor S_{ij} and those of four representative side chain bond vectors expressed as $\cos\phi_i^{C\beta C\gamma}\cos\phi_j^{C\beta C\gamma}$ in the sub-nanosecond timescale for four different residues of ubiquitin: (**a**) Val5 (**b**) Lys48 (**c**) Leu56 (**d**) Leu71

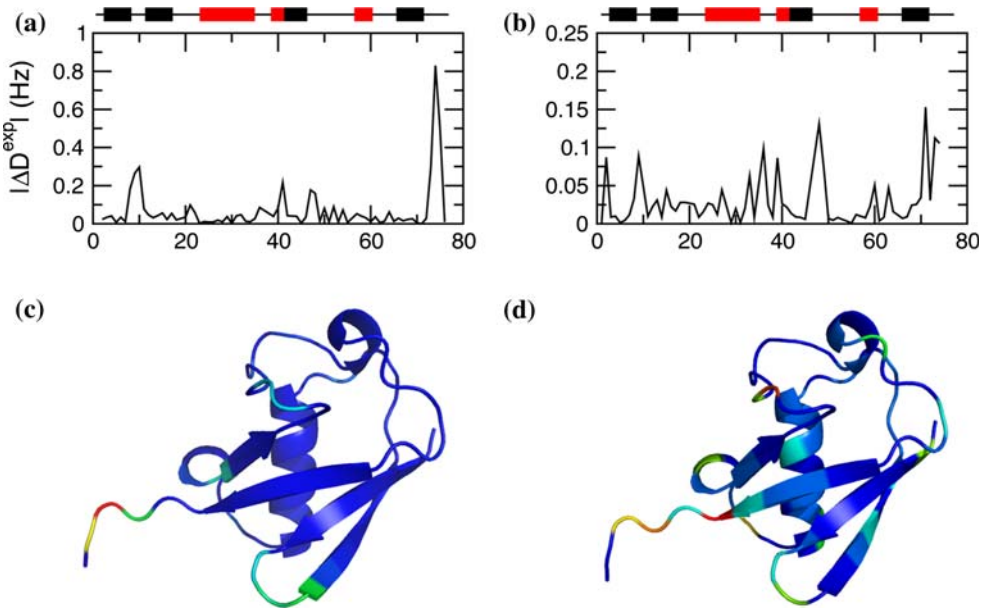

Fig. 8 Plot of ΔD^{exp} as a function of residue number (**a**) for the *NH* bond vectors. (**b**) for the $C_\beta C_\gamma$ bond vectors. Average structure of ubiquitin (PDB code 1d3z (Cornilescu et al. 1998)) colored according to the value of $|\Delta D^{exp}|$ of the *NH* (**c**) and $C_\beta C_\gamma$ (**d**) bond vector of each residue

correlations presented in Figs. 4–7 and the corresponding error $|\Delta D^{exp}|$. Nevertheless, it is clear that those residues for which the correlations are strongest present the largest deviations of D^{exp}. For the *NH* bond vectors these residues are Leu8 to Lys11, which define the turn between β-strands 1 and 2, residue Asp21 in the loop between β-strand 2 and α-helix 1, residue Gln41 in the first 3_{10}-helix, residues Gly47 and Lys48 in the loop between β-strand 3 and the second 3_{10}-helix and residues Leu73 to Gly75 at the C-terminus of the protein. These problematic residues are located in turns, loops and at the termini of the protein; indeed the average value of $|\Delta D^{exp}|$ for these residues is 0.12 Hz, whereas that of residues in elements of regular secondary structure is 0.04 Hz. The relationship between secondary structure and $|\Delta D^{exp}|$ becomes particularly clear by inspection of Fig. 8, where the residues of the average structure of ubiquitin have been colored according to the impact of correlations on the experimental *RDC*, D^{exp}.

There therefore exists a clear relationship between the correlations that are apparent in Fig. 6 and the small error in the *RDCs* presented in Fig. 8a, which is largest in the most dynamic regions of the protein.

The relationship between the location of the residue and the value of $|\Delta D^{exp}|$ is less conspicuous for the $C_\beta C_\gamma$ bond vector. The average value of $|\Delta D^{exp}|$ is 0.03 Hz both for residues in elements of secondary structure and for residues in turns, loops and at the termini; this result is not surprising given that also the degree of correlations present (Fig. 5) was less predictable for the $C_\beta C_\gamma$ bond vectors than it was for the *NH* bond vectors. This difference is caused by the fact that side chains are more dynamic than the protein backbone (Lindorff-Larsen et al. 2005), and that the motions of side chains are largely independent of their location in the protein structure (Best et al. 2004). A direct comparison of the average values of ΔD^{exp} obtained for the *NH* and $C_\beta C_\gamma$ bond vectors should take into account the dependence of the *RDCs* on the length of the internuclear vector and on the gyromagnetic ratios of the nuclei. It is possible to obtain the factor that enables us to rescale the side chain $C_\beta C_\gamma$ *RDCs* to the backbone *NH RDCs*

$$\frac{\gamma_N \gamma_H}{\gamma_C^2} \cdot \frac{r_{C_\beta C_\gamma}^3}{r_{NH}^3} \approx -3.6 \tag{8}$$

For comparison, for the values of $|\Delta D^{exp}|$ obtained for *NH* and $C_\beta C_\gamma$ *RDCs*, we found that the error in the average orientation of the side chains is approximately 2.5 times larger than that of the *NH* backbone vectors. This result is in agreement with the data presented in Fig. 5 that shows that the correlations between S_{ij} and $\cos\phi_i \cos\phi_j$ are more significant for the $C_\beta C_\gamma$ side chains than for the *NH* backbone bond vectors. Therefore the results in Fig. 8 demonstrate that the presence of correlations between the motions of S_{ij} and those of $\cos\phi_i \cos\phi_j$ does affect the outcome of methods for the determination of the average structure of proteins using information derived from *RDC* measurements.

The practical consequences of $|\Delta D^{exp}|$ for the refinement of the average structure of proteins using *RDCs* depends on its size, relative to the error δD^{exp} in the experimental measurement of the ensemble-averaged *RDC*. Refinement methods based on restrained molecular simulations (Schwieters et al. 2003) take into account the uncertainty δD^{exp} in the value of D^{exp} by adjustment of the force constant α used to restrain the position of the bond vector (Eq. 9) during refinement, using a quadratic potential E_{RDC}

$$E_{RDC} = \alpha \sum_i (D_i^{exp} - D_i^{calc})^2 \tag{9}$$

where i runs through the list of *RDC* restraints and it is assumed that D^{exp} has a Gaussian distribution centered around the measured value and with a standard deviation σ equal to the experimental error δD^{exp}.

The main sources of uncertainty in the measurement of D^{exp} are the intrinsic linewidth of the NMR signal and the digital resolution of NMR experiment used to measure the *RDCs*, which can be considered to a first approximation independent of the nature of the nuclei; a single value of α can thus be used in implementing Eq. 6. The information provided by a set of *RDCs* depends therefore on the value of the prefactor C_{PQ}; for pairs of nuclei with a high value of C_{PQ}, leading to large values of D^{exp}, such as for the *CH* bond vectors of aromatic side chains, the information content of the *RDCs* is high. By contrast, for the $C_\beta C_\gamma$ bond vectors, that have a low value of C_{PQ}, the *RDCs* contain less information. In brief, in the refinement of the average structure of proteins using multiple sets of *RDCs* in a given alignment medium, the set with the largest value of C_{PQ} dominates the refinement because α does not strongly depend of the nature of nuclei.

However, the presence of correlations between the fluctuations of $\cos\phi_i \cos\phi_j$ and S_{ij} introduces an additional source of uncertainty, ΔD^{exp}, on the ability of Eq. 4 to report on the exact distribution of structures that gives rise to the ensemble-averaged *RDC*. The error ΔD^{exp}, as opposed to the experimental uncertainty δD^{exp}, does depend on the nature of the bond vector i.e., is proportional to C_{PQ} for a given degree of correlation and is therefore, at least in principle, not optimally accounted for by current methods for the determination of the average structure of the protein from *RDCs*. As shown in Fig. 8 for bond vectors that yield low *RDCs*, such as the $C_\beta C_\gamma$ case studied in this work, ΔD^{exp} is lower than δD^{exp}, which is typically 0.5 Hz, thus only marginally increasing the total uncertainty on the value of D^{exp} and, as a consequence, of the average orientation of the bond vector with respect to the reference frame. By contrast, for bond vectors which yield large *RDCs* such as the *CH* bond vectors of aromatic side chains, the error ΔD^{exp} that will arise if correlations are present can become larger than the experimental error δD^{exp}; in this case the overall uncertainty in the position of the bond vector can increase significantly and lead to inaccuracies in the determination of the average structure.

In practice, the results that we have obtained show that the presence of tensor motions does not in general significantly affect the calculation of the average structure of proteins from *RDCs*, particularly for backbone bond vectors in elements of secondary structure. However, care needs to be exercised in the analysis of bond vectors that present significant correlations, which are generally found in the most dynamics regions of the protein, particularly when they have large values of C_{PQ}.

Effects of the correlations on the analysis of the dynamics of proteins using RDCs

We have shown how the presence of correlations can in principle affect the outcome of methods for the determination of the average structure of proteins. These correlations can also have an effect on the estimation of the amplitude of bond vector dynamics because the approaches that can be used for this type of study are based on the assumption of an absence of correlated dynamics in the alignment tensor. In establishing the size of this effect we have to consider that the order parameter, S, used to describe the motion of bond vectors, must be defined in a fixed molecular reference frame. It is therefore necessary to assume that the factorization used in Eq. 2 is correct before proceeding with an analysis. We thus compare the model-free order parameters S obtained from model-free approaches when the spherical harmonics are fit to the exact *RDCs*, $D^{\exp}$, to the order parameters S obtained when they are fit to the approximate, *RDCs,* $D^{\exp,*}$. It is unfortunately not possible to rigorously carry out such an analysis here because the numerical computation of the alignment tensor, which is necessary to calculate $D^{\exp}$ and $D^{\exp,*}$, is only available for a limited number of alignment mechanisms (Zweckstetter and Bax 2000; Zweckstetter et al. 2004), making it difficult to generate a sufficiently large number of independent sets of *RDCs* to determine the spherical harmonics and the order parameters with confidence.

However, it is still possible to compare the effect of correlations on the S order parameters derived from *RDCs* if the motions of the bond vectors are analyzed within a simple motional model. Because motional models make implicit assumptions about the motions that they allow, an approximate order parameter can be computed using a limited set of independent *RDCs*. Indeed, whereas five independent sets of *RDCs* are required for the determination of the spherical harmonics (Meiler et al. 2001) or for the analysis of bond vector motions using the 3D-GAF model (Bouvignies et al. 2005), only two are needed for 1D-GAF model (Bernadó and Blackledge 2004) and only one is needed for the simplest motional model, which assumes isotropic motion around the average position (Bax et al. 2001; Meiler et al. 2001). In order to study the influence of correlations on the analysis of the fluctuations of bond vectors, we computed the error introduced in the determination of S_{iso}, the S order parameter calculated using the isotropic motional model, because, due to its simplicity, it requires only one set of experimental *RDCs*. Although it is well established (Bernadó and Blackledge 2004) that the motions of bond vectors are often anisotropic, this approximation is very useful here because it allows us to connect, in principle, the correlations that we observe and the error that they introduce in the estimation of dynamics within this motional model, *i.e.* in terms of S_{iso}. The scaling factor relating the experimental *RDC*, $D^{\exp}$, to the *RDC* of the average structure, D^{av}, is equivalent to the square root of the order parameter commonly used in relaxation analysis when the motion of the bond vector is isotropic (Bax et al. 2001; Meiler et al. 2001), yielding

$$D^{\exp} = S_{iso} D^{av} \tag{10}$$

In order to study the effect of the motions of the alignment tensor on the estimation of bond vector dynamics from *RDCs*, it is therefore possible to compute the difference ΔS_{iso} between the order parameter obtained using the exact experimental *RDCs*, S_{iso}, and the order parameter, S^*_{iso}, obtained using the *RDCs* computed assuming an absence of correlations

$$\Delta S_{iso} = S_{iso} - S^*_{iso} = \frac{\Delta D^{\exp}}{D^{av}} \tag{11}$$

This equation shows that the effects of correlations on the estimation of the amplitude of the fluctuations of bond vectors in the isotropic motional model depend on both the presence of correlations, which lead to non-zero values of $\Delta D^{\exp}$, and the average direction of the bond vector with respect to the main direction of alignment. At orientations with θ (Eq. 2) close to the magic angle, very small values of D^{av} result, leading to large errors ΔS_{iso} if correlations are present i.e., if $\Delta D^{\exp} \neq 0$. Thus the sensitivity to the presence of correlations of methods for the analysis of the amplitude of protein dynamics is tuned by the actual value of the *RDC*. The results of the analysis for the *NH* and $C_\beta C_\gamma$ bond vectors are shown in Figs. 9 and 10, where the relationship between $\Delta D^{\exp}$ and ΔS_{iso} is very clear.

When only one set of *RDCs* is used for the refinement of the average structure of a protein, the information content of the *RDCs* of bond vectors that are parallel to the magic angle of the eigenframe of the alignment tensor is low because the corresponding value of $D^{\exp}$ can become comparable in size to the experimental error in the measurement of the *RDC* (that is independent of the identity of the bond vector) and to the error due to the presence of weak correlations, that is instead a function of the bond vector as described in the previous section. This problem can, in practice, be alleviated by using sets of *RDCs* measured in different independent alignment media; this procedure not only increases the amount of experimental information used to refine the average structure but also allows for the degree of refinement to become more homogeneous across the reference frame. Hence by using a combination of different alignment media the effect of experimental error and of the assumptions of the method can be greatly diminished.

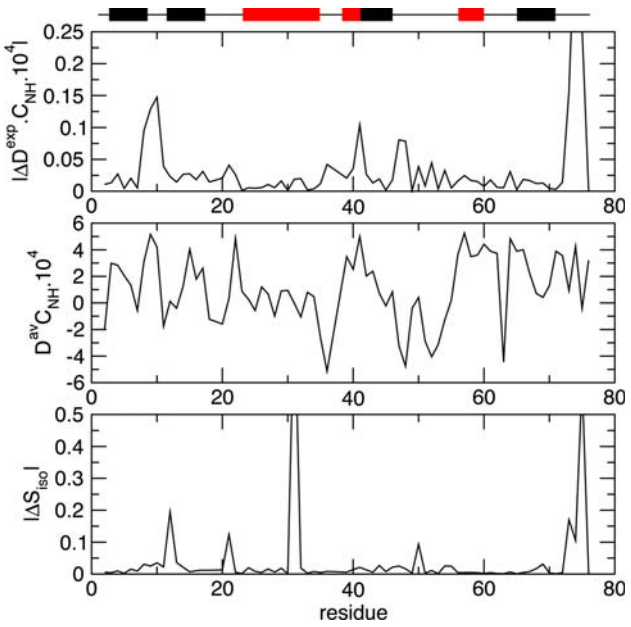

Fig. 9 (**a**) Plot of $|\Delta D^{exp}|$ and D^{av} as a function of residue number for the *NH* bond vectors of ubiquitin after normalization. (**b**) Plot of $|\Delta S_{iso}|$. In these plots the elements of secondary structure are shown as colored rectangles (β-strands are shown in black and α-helices are shown in red)

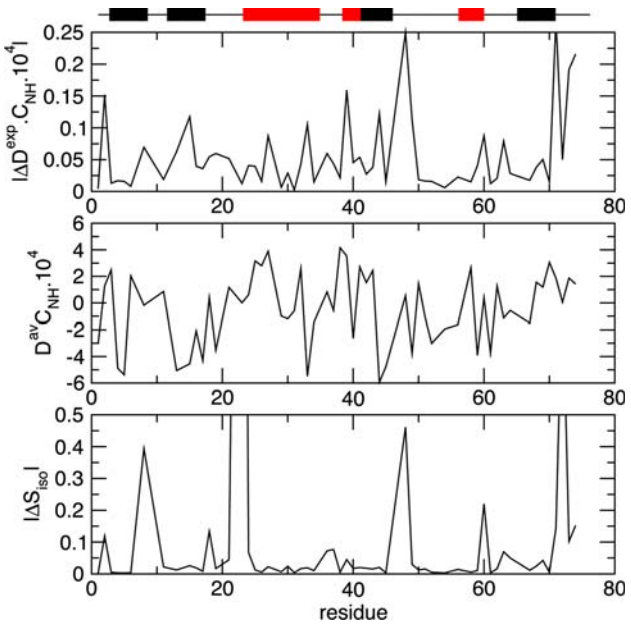

Fig. 10 (**a**) Plot of $|\Delta D^{exp}|$ and D^{av} as a function of residue number for the $C_\beta C_\gamma$ bond vectors of ubiquitin after normalization. (**b**) Plot of $|\Delta S_{iso}|$. In these plots the elements of secondary structure are shown as colored rectangles (β-strands are shown in black and α-helices are shown in red)

The results that we present here suggest that a similar situation is encountered in the study of protein dynamics using *RDCs*, where the difference between the *RDC* of the average structure, D^{av}, and that measured experimentally, D^{exp}, is analyzed to obtain information about the fluctuations of the bond vectors of the protein. The analysis of protein dynamics is in fact more sensitive to the value of the *RDC* due to the relationship between the amplitude of the fluctuation, expressed as S_{iso} within a simple motional model (Eq. 8), and to the value of *RDC* of the average structure, D^{av}. Since the use of several alignment media, at least five, is necessary for the analysis of protein dynamics at high resolution (Meiler et al. 2001; Bouvignies et al. 2005) we suggest that the effects of the weak correlations that we observe do not significantly affect the results of these analysis. We envisage two mechanisms by which the use of several alignment mechanisms can alleviate the problems that we present in Figs. 9 and 10; on the one hand, the identity and strength of correlations will almost certainly be alignment-specific and on the other hand the value of the *RDC* for a given bond vector in the different alignment media will vary.

However, the results that we present in this work unequivocally show that even weak correlations can have significant effects in the analysis of the structure and dynamics of a relatively rigid protein, and thus highlight the need to develop methods where the fluctuations of the alignment tensor of the protein are explicitly taken into account. These methods will be particularly important for the study of the functional large scale motions that take place in the μs to ms timescale, such as interdomain motions, that can be in principle be characterized at high resolution using *RDCs*.

Conclusions

Although the steric alignment tensor of the protein ubiquitin experiences very significant fluctuations in the sub-nanosecond timescale, we have shown that for residues in secondary structure elements these fluctuations do not significantly affect the extraction of backbone dynamics using steric *RDCs* because they are largely uncorrelated with the motions of the relevant bond vectors. We have also found, however, that the analysis of side chain motions is often affected by weak correlations that are difficult to predict from the average structure of the protein, at least in a steric alignment medium. The effects of such correlations are small for sub-nanosecond fluctuations in native proteins such as ubiquitin, because in this case the dynamics of the protein can be very accurately described in terms of structural fluctuations around an average structure. However, when the protein experiences large-scale cooperative biologically relevant motions, which include interdomain motions and large scale conformational changes, it will be important to develop methods that explicitly take into account the fluctuations of the alignment tensor.

Acknowledgments The authors would like to thank Nils-Alexander Lakomek, Edward d'Auvergne, Donghan Lee and Christian Griesinger for very helpful discussions. This work was supported by grants from the Leverhulme Trust (XS and MV), the National Science Foundation (BR) and the Royal Society (MV).

References

Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Prot Sci 12:1–16

Bax A, Grishaev A (2005) Weak alignment NMR: a hawk-eyed view of biomolecular structure. Curr Opin Struct Biol 15:1–8

Bax A, Kontaxis G, Tjandra N (2001) Dipolar couplings in macromolecular structure determination. Meth Enzymol 339: 127–174

Bernadó P, Blackledge M (2004) Local dynamic amplitudes on the protein backbone from dipolar couplings: toward the elucidation of slower motions in biomolecules. J Am Chem Soc 126: 7760–7761

Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci USA 102:17002–17007

Best RB, Clarke J, Karplus M (2004) The origin of protein side-chain order parameter distribution. J Am Chem Soc 126:7734–7735

Blackledge M. (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. Prog Nucl Mag Res Spec 46:23–61

Bouvignies G, Bernadó P, Meier S, Cho K, Grzesiek S, Bruschweiler R, Blackledge M (2005) Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. Proc Natl Acad Sci USA 102:13885–13890

Bouvignies G, Marwick P, Bruschweiler R, Blackledge M (2006) Simultaneous determination of protein backbone structure and dynamics from residual dipolar couplings. J Am Chem Soc 128:15100–15101

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization and dynamics calculations. J Comp Chem 4:187–217

Clore GM, Schwieters CD (2004a) How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? J Am Chem Soc 126:2923–2938

Clore GM, Schwieters CD (2004b) Amplitudes of protein backbone dynamics and correlated motions in a small α/β protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. Biochemistry 43:10678–10691

Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120:6836–6837

de Alba E, Tjandra N (2002) NMR dipolar couplings for the structure determination of biopolymers in solution. Prog Nucl Mag Res Spec 40:175–197

Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. Proc Natl Acad Sci USA 102:13099–13104

Lakomek N, Carlomagno T, Becker S, Griesinger C, Meiler J (2006) A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. J Biomol NMR 34:101–115

Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M (2005) Simultaneous determination of protein structure and dynamics. Nature 433:128–132

Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic relaxation rates in macromolecules. 1. Theory and range of validity. J Am Chem Soc 104:4546-4559

Louhivuori M, Otten R, Lindorff-Larsen K, Annila A (2006) Conformational fluctuations affect protein alignment in dilute liquid crystal media. J Am Chem Soc 128:4371–4376

Meiler J, Peti W, Griesinger C (2003) Dipolar couplings in multiple alignments suggest α helical motion in ubiquitin. J Am Chem Soc 125:8072–8073

Meiler J, Prompers JJ, Peti W, Griesinger C, Bruschweiler R (2001) Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. J Am Chem Soc 123:6098–6107

Peti W, Meiler J, Bruschweiler R, Griesinger C (2002) Model-free analysis of protein backbone motion from residual dipolar couplings. J Am Chem Soc 124:5822–5833

Richter B, Gsponer J, Várnai P, Salvatella X, Vendruscolo M (2007) The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. J Biomol NMR 37:117–135

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Res 160:66–74

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

Tolman J, Ruan K (2006) NMR residual dipolar couplings as probes of biomolecular dynamics. Chem Rev 106:1720–1736

Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1997) NMR evidence for slow collective motions in cyanometmyoglobin. Nat Struct Biol 4:292–297

Tolman JR, Al-Hashimi HM, Kay LE, Prestegard JH (2001) Structural and dynamic analysis of residual dipolar coupling data for proteins. J Am Chem Soc 123:1416–1424

Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 Å resolution. J Mol Biol 194:531–544

Zweckstetter M, Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J Am Chem Soc 122: 3791–3792

Zweckstetter M, Bax A (2002) Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. J Biomol NMR 23: 127–137

Zweckstetter M, Hummer F, Bax A (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. Biophys J 86:2039–2046