J Biomol NMR (2012) 53:281–292
DOI 10.1007/s10858-012-9644-3

ARTICLE

Determination of structural fluctuations of proteins from structure-based calculations of residual dipolar couplings

Rinaldo W. Montalvao · Alfonso De Simone · Michele Vendruscolo

Received: 5 October 2011 / Accepted: 19 May 2012 / Published online: 23 June 2012

Abstract Residual dipolar couplings (RDCs) have the potential of providing detailed information about the conformational fluctuations of proteins. It is very challenging, however, to extract such information because of the complex relationship between RDCs and protein structures. A promising approach to decode this relationship involves structure-based calculations of the alignment tensors of protein conformations. By implementing this strategy to generate structural restraints in molecular dynamics simulations we show that it is possible to extract effectively the information provided by RDCs about the conformational fluctuations in the native states of proteins. The approach that we present can be used in a wide range of alignment media, including Pf1, charged bicelles and gels. The accuracy of the method is demonstrated by the analysis of the Q factors for RDCs not used as restraints in the calculations, which are significantly lower than those corresponding to existing high-resolution structures and structural ensembles, hence showing that we capture effectively the contributions to RDCs from conformational fluctuations.

Keywords Protein dynamics · Protein conformational ensembles · Molecular dynamics simulations

Electronic supplementary material The online version of this article (doi:10.1007/s10858-012-9644-3) contains supplementary material, which is available to authorized users.

R. W. Montalvao · A. De Simone · M. Vendruscolo (✉)
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
e-mail: mv245@cam.ac.uk

A. De Simone
Division of Molecular Biosciences, Imperial College, South Kensington Campus, London SW7 2AZ, UK

Introduction

Native states of proteins are characterised by the presence of conformational fluctuations that are important for enabling their functions (Frauenfelder et al. 1991; Palmer 2004; Boehr et al. 2006; Mittermaier and Kay 2006; Tolman and Ruan 2006; Vendruscolo and Dobson 2006; Shaw et al. 2010; Kalodimos 2011; Salmon et al. 2011). It is therefore particularly important to develop techniques capable of providing accurate representations of these motions. Nuclear magnetic resonance (NMR) spectroscopy represents a powerful technique for this purpose (Palmer 2004; Boehr et al. 2006; Mittermaier and Kay 2006; Tolman and Ruan 2006; Vendruscolo and Dobson 2006; Kalodimos 2011; Salmon et al. 2011) since NMR measurements provide time- and ensemble-averaged results at atomic-level resolution.

Great attention has been devoted to the development of methods for characterising the conformational fluctuations of proteins and nucleic acids from NMR parameters, in particular PRE-derived distances (Iwahara et al. 2004; Bertoncini et al. 2005; Dedmon et al. 2005; Allison et al. 2009; Clore and Iwahara 2009; Huang and Grzesiek 2010), S^2 order parameters (Zhang and Bruschweiler 2002; Best and Vendruscolo 2004; Lindorff-Larsen et al. 2005; Richter et al. 2007), and residual dipolar couplings (RDCs) (Tjandra and Bax 1997; Tolman et al. 1997; Meiler et al. 2001; Bax 2003; Clore and Schwieters 2004a, b; Bouvignies et al. 2006; Iwahara et al. 2006; Tolman and Ruan 2006; Zhang et al. 2006; Showalter and Bruschweiler 2007; Zhang et al. 2007; Lange et al. 2008; De Simone et al. 2009; Huang and Grzesiek 2010; Fenwick et al. 2011). An aspect that requires particular attention in these approaches is that NMR parameters can be analysed only in an approximate manner in terms of individual structures, because of the intrinsic dynamical nature of proteins. In practice, when the

structural fluctuations are of limited amplitude, it is possible to identify an average structure that corresponds quite well to a given set of measurements. But if the structural fluctuations are more significant, it becomes necessary to represent the state of a protein by using an ensemble of structures, so that the average values of the NMR parameters over the ensemble reproduce closely the experimentally measured values. In this case, individual structures are not expected to exhibit values for the NMR parameters that match exactly the experimental ones. In this view, the problem of comparing the values of NMR parameters calculated from individual structures with those measured experimentally is not well defined, since the latter are resulting from a time and conformational averaging procedure that unavoidably takes place during the measurements. In the presence of significant structural fluctuations, therefore, computational methods for characterising the behaviour of proteins should be built around the comparison between experimental parameters and average values estimated over ensembles of conformations (Best and Vendruscolo 2004; Clore and Schwieters 2004a, b).

This problem is particularly relevant in the case of residual dipolar couplings (RDCs), since the values of these NMR parameters tend to have a strong conformational dependence, and hence they fluctuate significantly during the motion of a protein (Louhivuori et al. 2006; Salvatella et al. 2008). In practical terms, in the presence of conformational fluctuations of large amplitude even the most accurate methods for calculating the RDCs for a given structure (Zweckstetter and Bax 2000; Fernandes et al. 2001; Almond and Axelsen 2002; Azurmendi and Bush 2002; van Lune et al. 2002; Ferrarini 2003; Zweckstetter 2008; Berlin et al. 2009) may not provide values that match exactly the experimental ones. A close agreement between calculated and experimental RDCs can therefore be obtained only when the calculated RDCs are averaged over an ensemble of structures representing the motions of the protein (Clore and Schwieters 2004a, b; Showalter and Bruschweiler 2007; De Simone et al. 2009; Huang and Grzesiek 2010).

In this work we describe a method for using RDCs to characterise conformational fluctuations of proteins and illustrate it by determining structural ensembles representing the native state dynamics of two proteins, ubiquitin and acylphosphatase from *Sulfolobus solfataricus* (Sso AcP). The method that we discuss exploits the information provided by RDCs by incorporating them as structural restraints in molecular dynamics simulations. In order to take account of the conformational dependence of the RDCs, these parameters are calculated from the atomic coordinates of individual structures using a structure-based approach similar to PALES (Zweckstetter and Bax 2000). To demonstrate that this approach is capable of extracting in an effective manner the information about conformational fluctuations contained in RDC measurements, we consider the RDCs not used as restraints and show that the corresponding Q factors calculated from the ensembles of structures that we generated are significantly lower than those obtained from high-resolution individual structures.

Methods

Residual dipolar couplings

The RDC between two nuclear spins of gyromagnetic ratios γ_1 and γ_2 at a given distance r can be expressed as (Bax 2003; Blackledge 2005)

$$D = D_{\max}\left\langle \frac{1}{2}\left(3\cos^2\,\vartheta - 1\right)\right\rangle \tag{1}$$

where ϑ is the angle between the internuclear vector and the external magnetic field, $D_{\max} = -\mu_0\gamma_1\gamma_2 h/8\pi^3 r^3$ is the maximal value of the dipolar coupling for the two nuclear spins and μ_0 is the magnetic constant; the averaging specified by the square brackets is carried over the possible orientations of the internuclear vector with respect to the external magnetic field. In isotropic solutions RDCs average to zero because all directions are equivalent, but in an anisotropic medium the appearance of a preferred direction results in non-zero values of the RDCs.

In NMR studies of proteins it is convenient to introduce a 3×3 matrix A that specifies the orientation of the molecular reference frame with respect to the laboratory reference frame

$$A_{ij} = \frac{1}{2}\left(3\cos\psi_i\cos\psi_j - \delta_{ij}\right) \tag{2}$$

where $i, j = (x, y, z)$, δ_{ij} is the Kronecker delta function, and ψ_i is the angle between the ith molecular axis and the z axis in the laboratory frame. Using the matrix A, instead of using Eq. 1, the RDC of any given internuclear vector in the protein can be expressed as

$$D = D_{\max}\sum_{ij}\left\langle A_{ij}\cos\varphi_i\cos\varphi_j\right\rangle \tag{3}$$

where φ_i and φ_j are the angles between the internuclear vector and the molecular reference frame, and as above the averaging specified by the square brackets is carried over the possible orientations of the internuclear vector with respect to the external magnetic field.

Structure-based calculation of the alignment tensor

It is convenient to introduce the alignment tensor, which is denoted by $\langle A\rangle$, for a protein conformation, so one can recast Eq. 3 as

$$D = D_{\max} \sum_{ij} \langle A_{ij} \rangle \cos \varphi_i \cos \varphi_j \quad (4)$$

The alignment tensor is given through an averaging procedure that considers all the translations and rotations (R, Ω) of the protein conformation with respect to an alignment medium, where R defines the distance between centres of mass of the protein molecule and the closest molecule of the alignment medium and $\Omega = (\psi_1, \psi_2, \psi_3)$ defines the orientation of the protein molecule. The expression for the individual components of the alignment tensor is

$$\langle A_{ij} \rangle = \frac{1}{Z} \iint A_{ij}(r, \Omega) P_B(R, \Omega) dR \, d\Omega \quad (5)$$

where

$$Z = \iint P_B(R, \Omega) dR \, d\Omega \quad (6)$$

is the partition function. If the interaction between the protein molecule and the alignment medium is purely steric, the form of the Boltzmann factor P_B (R, Ω) is particularly simple, being zero if the protein is clashing with the alignment medium, and unity otherwise. If the alignment medium is electrostatically charged, for a particular configuration (R, Ω), the Boltzmann factor should be modified to include the electrostatic potential energy due to the interactions between the protein molecule and the alignment medium, which is

$$E_{elec}(R, \Omega) = \sum_i q_i \psi[r_i(R, \Omega)] \quad (7)$$

where the partial charges q_i for the atoms at positions r_i were obtained from the Biochemical Algorithms Library (Hildebrandt et al. 2010). In this case P_B (R, Ω) is given by the Boltzmann factor

$$P_B(R, \Omega) = \frac{1}{Z} \exp\left[-\frac{E_{elec}(R, \Omega)}{k_B T} \right] \quad (8)$$

where T is the temperature and k_B is the Boltzmann constant. Finally, the alignment tensor $\langle A \rangle$ needs to be multiplied by the order parameter of the alignment medium due to its imperfect alignment with the magnetic field.

We illustrate the process for calculating the alignment tensor through Eq. 5 in the case of bicelles or gels, which are commonly used alignment media. The calculation starts by centring the protein at the origin of the coordinate system and determining its surface by projecting a spherical grid on to the exposed atoms. This grid is composed of $51^2 = 2{,}601$ points evenly distributed on a sphere of 150 Å radius. The projection is done by tracing rays that starts from grid points and move towards the origin of the system until stopped by one of the balls that represent the atoms of the protein. The radius of each ball is chosen as the van der Waals radius of the corresponding atom, which is taken from the Biochemical Algorithms Library (Hildebrandt et al. 2010). This process produces a list of points on the surface that are likely to collide with the bicelles during the steric sampling without spending computational time in dealing with too many surface details. The next step is the rotational and radial sampling of the sterically allowed orientations and it is composed of two distinct parts. First, the superposition of the z axis of the molecular frame into the axis defined by the origin and one point on the unit sphere and the rotation around this axis in steps of 20°. These two steps are repeated for $15^2 = 225$ points evenly distributed on the unity sphere, thus producing a total of 4,050 orientations for the protein. Second, for each orientation, the occurrence of steric clashes between the protein and the bicelles is tested by positioning an infinite wall, parallel to the z axis, in a one-dimensional grid whose points are spaced by 0.5 Å and than checking if any of the surface points are beyond the limit imposed by this wall. The radial sampling is limited to a maximum distance where there are no steric clashes and the electrostatic potential is very weak.

We adopted the Gouy-Chapmman theory (Zweckstetter et al. 2004), and assumed that the bicelles or the gels can be represented by an infinite flat wall with a charge density σ^0 in contact with a fluid with symmetrical electrolytes of charge density ρ. In this model, ions are considered point charges, the ionic adsorption energy is purely electrostatic, the average electrostatic potential is identified with the potential of mean force, the solvent is represented by a structureless continuum with a constant dieletric permittivity ε. For an infinite flat wall in the ($x = 0$, y, z) plane, the electrostatic potential ψ is only dependent on the distance from the wall to the considered position in the x axis; and the analytical solution for Poisson–Boltzmann equation, in the absence of the protein, is

$$\frac{e}{k_B T} \psi(x) = \frac{4}{z} \left\{ \left(\tanh\left[\frac{1}{2} \sinh^{-1}\left(\frac{e\sigma^0}{2 k_B T \varepsilon_0 \varepsilon \kappa} \right) \right] \right) e^{-\kappa x} \right\} \quad (9)$$

where ε_0 is the vacuum permittivity, ε the relative dielectric constant, and e the electron charge. In this equation κ is the inverse Debye length

$$\kappa^2 = \frac{2 e^2 z^2 \rho}{\varepsilon_0 \varepsilon k_B T} \quad (10)$$

This solution is only valid for symmetrical electrolytes (where $z_+ = z_- = z$), as the Poisson–Boltzmann equation for asymmetrical electrolytes or mixed electrolytes has no known analytical solution even for such a simple case as of a charged infinite flat wall. In practice, the absence of an

analytical solution could constitute a problem as experiments for measuring RDCs often employs sodium phosphate that is an asymmetrical electrolyte under the standard experimental conditions. The combination sodium phosphate with sodium chloride, which creates a system with mixed-type electrolyte, is also very common. However, the symmetrical model can be a very good approximation for both systems for low electrolyte concentrations (mM).

In its current implementation, the method discussed here calculates the alignment tensor for neutral and charged bicelles, polyacrylamide gels and bacteriophage Pf1. For modelling the electrostatic potential in the case of charged Pf1 media, we employ a cylindrical double layer model (Ferrarini 2003) with symmetric monovalent unstructured ions and asymptotically vanishing potential. The cylinder is modelled as an infinite dielectric rod possessing a constant charge density (-5.0×10^{18} C/m^2) and the Poisson–Boltzmann equation is solved numerically in the absence of Pf1 under Neumann conditions (Zweckstetter et al. 2004).

Averaging of the RDCs over the positions of the electrostatic charges

A challenge in the calculation of RDCs for a given structure is created by the very high sensitivity that these NMR parameters have on the thermodynamic conditions (Zweckstetter et al. 2004). Since the RDCs are provided by an averaging procedure over the possible orientations of a protein molecule, and the probabilities of such orientations depend strongly on electrostatic interactions, it is difficult to perform accurate calculations. The Boltzmann weight corresponding to a given orientation depends on the temperature, the pH, the ionic strength and the distribution of charges on the protein, which are all factors that can currently be calculated only within some error. As a result, the averaging procedure over the different orientations of a protein molecule, which provides its RDCs, results in values that are often rather approximate. In order to take account of this problem we perform a procedure in which the values of the RDCs are obtained by averaging over the distributions in the positions of each charge. These distributions are obtained by considering ten conformations taken at 0.5 ps intervals from a 5 ps window in the molecular dynamics trajectory of each replica. The specific time intervals and number of conformations considered were chosen in order to optimise the agreement between experimental and calculated RDC values, as assessed by the Q factors (Eq. 14), in a series of short preliminary molecular dynamics simulations carried out starting from the X-ray structure of Sso AcP.

Molecular dynamics simulations with replica-averaged RDC restraints: the SVD method

In this case, molecular dynamics simulations with ensemble-averaged RDC restraints were implemented in the Gromacs package (Hess et al. 2008) and the Amber99SB (Hornak et al. 2006), by adopting the SVD calculation of the alignment tensor originally described by Clore and Schwieters (2004a, b), De Simone et al. (2009), Fenwick et al. (2011). In this method, the alignment tensor of a given structure is fitted using the SVD method to the experimentally measured RDCs. A single tensor orientation was used for all the members of the ensemble, but each ensemble member was allowed to take its own values of the axial and rhombic components to generate alignment tensors consistent with ensemble members with different shapes. The values of the axial and rhombic components of the different ensemble members where allowed to differ by only a small amount, by introducing energy terms to restrain their spread using parameters described previously (Clore and Schwieters 2004a, b; De Simone et al. 2009; Fenwick et al. 2011).

Molecular dynamics simulations with replica-averaged RDC restraints: the SB method

In this case, molecular dynamics simulations with ensemble-averaged RDC restraints were implemented in the Gromacs package (Hess et al. 2008), by adopting the structure-based calculation of the alignment tensor described by Eqs. (5–10). In this approach, restraints are imposed by adding a restraint term, E_{RDC}, to a standard molecular mechanics force field, E_{MM},

$$E_{TOT} = E_{MM} + E_{RDC} \tag{11}$$

The resulting force field, E_{TOT}, was employed to integrate the equations of motion. The E_{MM} that we employed was the Amber99SB (Hornak et al. 2006) and the restraint term is given by (De Simone et al. 2009)

$$E_{RDC} = \alpha \sum_i \left(D^{exp} - D^{calc}\right)^2 \tag{12}$$

where α is the weight of the restraint term, and D^{exp} and D^{calc} are the experimental and calculated RDCs, respectively. For a given bond vector PQ, the calculated RDC is evaluated as

$$D^{calc} = \frac{1}{M} \sum_{m=1}^{M} D_m \tag{13}$$

where m runs over the M replicas and D_m is the RDC of replica M, which is given by Eq. 4. The choice of M is dictated by the trade-off between the two opposite effects

of overfitting and over-restraining (Richter et al. 2007; Vendruscolo 2007). By using large values of M one increases the number of effective parameters available to fit the experimental data. However, this strategy can incur in the problem of overfitting, since the structures obtained from the simulations will not be consistent with experimental data not used as restraints (Richter et al. 2007; Vendruscolo 2007). By using one or too few replicas, one has instead the problem of over-restraining, at least in cases in which the conformational fluctuations are of large amplitude, as mentioned in the Introduction (Richter et al. 2007; Vendruscolo 2007).

The sampling of the conformational space is preceded by an equilibration simulation at 300 K in which the restraints are gradually introduced by raising the force constant α from zero to a value set on the basis of the level of agreement with experimental data in order to provide the best agreement between calculated and experimental data (Eq. 12). Since the alignment tensor is calculated directly from the structures (Eq. 5), it is rescaled by a global factor corresponding to the inverse of the slope of the correlation between experimental and calculated data. In the sampling phase, a series of 20 cycles of simulated annealing between 300 and 500 K is carried out to increase the efficiency in the sampling of the conformational space. During these cycles, the force constant for the restraints is also annealed so that at high temperatures the force is tenfold higher. Each cycle is carried out for a total of 250 ps (125,000 molecular dynamics steps) by using an integration step of 2 fs.

In the restrained simulation technique used here, the restraints are imposed as averages over M replicas of the protein molecule. In the case of ubiquitin, we have previously shown that the optimal number of replicas is $M = 8$ (De Simone et al. 2009), which is also the number used in the present study. A total of 8,000 conformations were collected by extracting them from the part of the annealing cycle at 300 K. We then randomly selected from these conformations an ensemble of 160 structures for deposition in the PDB (code 2LJ5), and checked that this small ensemble was representative of the complete one in terms of standard structural and NMR parameters.

No further experimental information was used in the calculations in addition to the RDC restraints described above.

Summary of the computational approach

In this work we describe the method of using RDCs as structural restraints in molecular dynamics simulations that is summarized in Fig. 1:

(1) The integration of the equations of motion generates a sampling of conformational space that results in an ensemble of conformations (Fig. 1, pink rectangle) that can be used to calculate the properties of interest of a protein.
(2) In order to extract the information from the experimentally measured RDC values, the equations of motion contain a term (Eq. 12) that at each time step penalizes deviations between experimental and calculated RDC values (Fig. 1, blue rectangle).
(3) Since the experimental RDC values results from time and ensemble averages over the protein molecules in test tube, the calculated RDCs are averaged over multiple replicas of the protein molecule (Eq. 13; Fig. 1, blue rectangle).
(4) In order to calculate the RDCs we determine the alignment tensor from the shape and charge of the instantaneous conformation of each replica (Eqs. 1–5 and Fig. 1, green circle).
(5) Since the alignment tensor is very sensitive to the fast timescale fluctuations of the electrostatic charges, a preliminary averaging is carried out over the positions of the charges of each molecule (Fig. 1, red circle).

Summary of the different averaging procedures described in this work

In the approach that we describe in this work we use four different types of averaging procedures. The first is an averaging for each individual conformation over the small-amplitude fluctuations in the positions of the electrostatic charges (Fig. 1, red circle). The second is the averaging over the relative positions (in terms of rotations and translation) of the protein with respect to the alignment medium (Eqs. 1–5; Fig. 1, green circle). The third is the averaging at each time step over the replica used in the molecular dynamics simulations (Eq. 13; Fig. 1, blue rectangle). The fourth is the averaging over the internal dynamics of the protein, which in our approach is carried out at the end of the molecular dynamics simulations over all the conformations generated (Fig. 1, pink rectangle).

Experimental measurements of RDCs for SsoAcP

In order to obtain information for the determination of the native state ensemble of Sso AcP we carried out measurements of N-HN RDCs in bacteriophage Pf1 at 200 mM NaCl (Table S2).

Results

In this work we describe an approach in which RDC measurements are implemented as replica-averaged structural restraints in molecular dynamics simulations in order

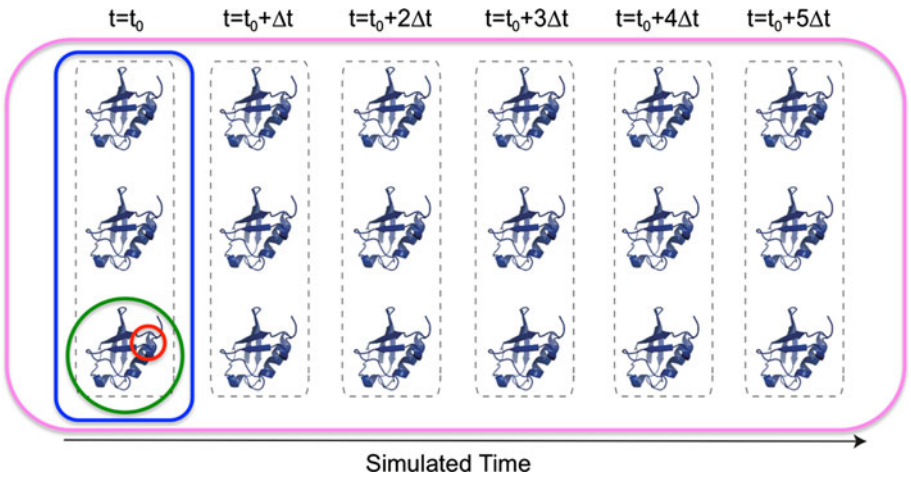


Fig. 1 Scheme of the computational method described in this work to generate a sampling of the conformation space of a protein molecule (*pink rectangle*). Molecular dynamics simulations are carried out in parallel for multiple copies (*replicas*) of the protein molecule. At each time step a biasing force is introduced to minimise the differences between experimental and calculated RDCs, which are averaged over the replicas (*blue rectangle*). The calculation of the RDCs for each individual replica at each time step is carried out by determining the alignment tensor from the shape and charge of the replica (*green circle*). The effects of the fast timescale fluctuations in the positions of the electrostatic charges are averaged out (*red circle*) before the calculation of the alignment tensor

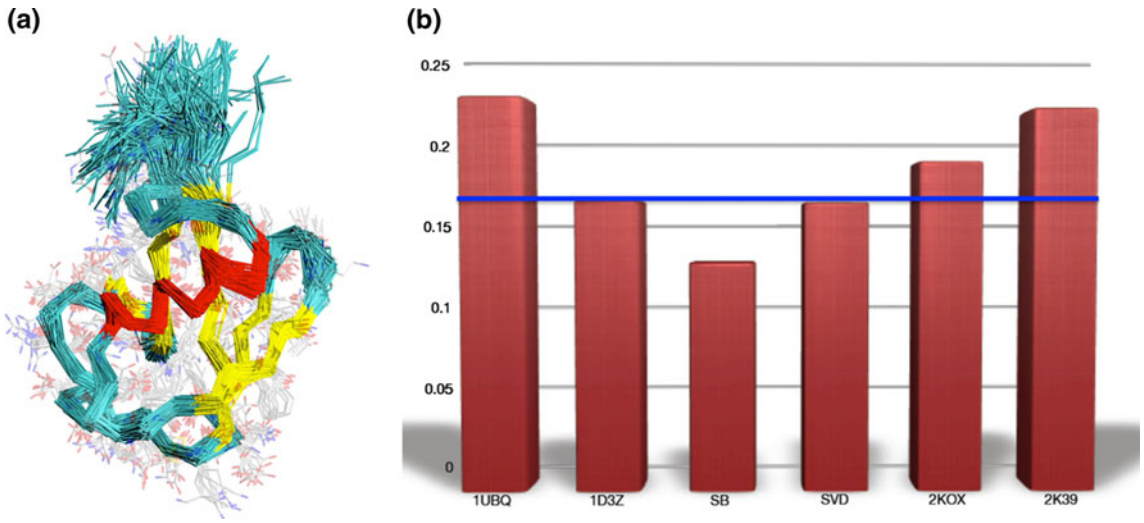


Fig. 2 **a** Ensemble of structures of ubiquitin (PDB code 2LJ5) derived from structure-based RDC restraints as described in this work (the 'SB 12 sets' ensemble). **b** Comparison of the Q factors for RDCs corresponding to various structures and structural ensembles of ubiquitin (Table 1): (1UBQ) X-ray structure (Vijay-Kumar et al. 1987), (1D3Z) RDC-based NMR structure (Cornilescu et al. 1998), (SB) RDC-based NMR ensemble with structure-based calculation of the alignment tensor, (SVD) RDC-based NMR ensemble with SVD-based calculation of the alignment tensor, (2KOX) RDC-based NMR ensemble with SVD-based calculations of the alignment tensor (Fenwick et al. 2011), (2K39) RDC-based NMR ensemble (Lange et al. 2008). The blue horizontal line indicates the Q factor for the 1D3Z, which is a representation of the average structure of ubiquitin, but not of its conformational fluctuations

to characterise the native state conformational fluctuations of proteins. In order to illustrate this approach we consider the case of two proteins, ubiquitin and Sso AcP.

Determination of the structural fluctuations in the native state of ubiquitin

By calculating the RDCs using the structure-based (SB) method described in this work (see "Methods"), we generated an ensemble of structures (SB '12 sets' ensemble, PDB code 2LJ5, Fig. 2a; Figures S1–S9) to represent the conformational fluctuations of this protein. As restraints we used 12 sets of RDCs, and for validation we used another 38 sets of RDCs (Table S1). In order to determine whether the approach that we followed is capable of extracting more information than other available ones we compared the SB ensemble with a X-ray structure of ubiquitin (1UBQ) (Vijay-Kumar et al. 1987), an NMR

Table 1 Comparison of the Q factors [Eq. (14)] corresponding to various structures and structural ensembles of ubiquitin: (1UBQ) X-ray structure (Vijay-Kumar et al. 1987), (1D3Z) RDC-based NMR structure (Cornilescu et al. 1998), (SB 12 sets, 2LJ5; SB 4 sets) RDC-based NMR ensembles with structure-based calculation of the alignment tensor, (SVD 12 sets) RDC-based NMR ensemble with SVD-based calculation of the alignment tensor, (2KOX) RDC-based NMR ensemble with SVD-based calculations of the alignment tensor (Fenwick et al. 2011), (2K39) RDC-based NMR ensemble (Lange et al. 2008)

	1UBQ	1D3Z	SB 12 sets	SB 4 sets	SVD 12 sets	2K39	2KOX
Q (restrained)	0.225	0.127	0.118	0.122	0.156	0.210	0.169
Q (unrestrained)	0.220	0.175	0.132	0.150	0.165	0.217	0.191
Q (total)	0.222	0.164	0.129	0.148	0.163	0.216	0.186

Restrained Q factors are calculated using the SVD method over the alignment media used as restraints in the 'SB 12 sets' case (Table S1) and unrestrained Q factors over the alignment media not used as restraints in the same case (Table S1). As these different RDC sets have different experimental precision the simple averaging of the Q factors used here provides an approximate indication of the quality of the structures

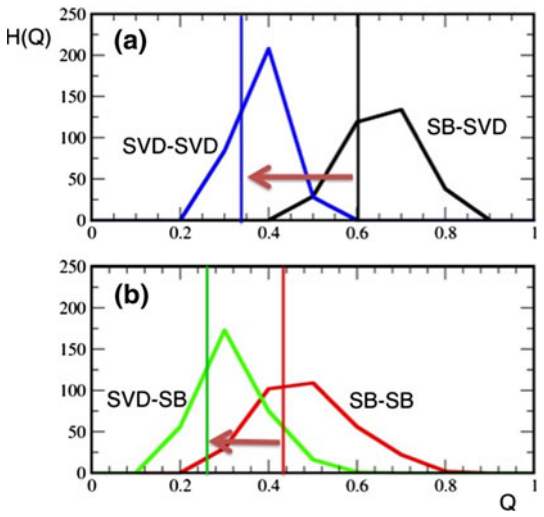


Fig. 3 Comparison of the Q factor distributions for the SVD and SB ensembles for ubiquitin. **a** SVD-ensemble. For each structure in the SVD ensemble we used both the SVD and the SB methods for back-calculating the RDCs to calculate the Q factors, Eq. (14). The Q factor distribution when the RDCs are back-calculated using the SB method is indicated as SB-SVD, and as SVD–SVD when the RDCs are back-calculated using the SVD method (see also Table 2). The arrow indicates the extent of overfitting in the SVD back-calculation as the difference between the average Q factors (represented as *vertical lines*) calculated using the SVD and the SB back-calculations. **b** Analogous results for the SB-ensemble (see also Table 2)

structure determined using RDCs (1D3Z) (Cornilescu et al. 1998) and an ensemble of structures (SVD '12 sets' ensemble) obtained by using a method similar to the one discussed in this paper but in which the alignment tensor is not calculated from the structures, but by a fitting procedure, the singular value decomposion (SVD) method (Losonczi et al. 1999). The results (Fig. 2b; Table 1) indicate that the SVD ensemble, at least in the implementation that we adopted here (Clore and Schwieters 2004a, b; De Simone et al. 2009; Fenwick et al. 2011), does not capture the dynamics of ubiquitin in a highly accurate manner. This conclusion is obtained by considering the Q factors for RDCs. The Q factor is defined as (Bax 2003)

Table 2 Comparison of the average Q factors [denoted as $\bar{Q}$, Eq. (15)] calculated for individual structures in the ensembles determined for ubiquitin (column 3) and Sso AcP (column 4)

Back-calculation	Restraints	Ubiquitin	Sso AcP
SB	SVD	0.61	0.70
SVD	SVD	0.34	0.36
SB	SB	0.44	0.55
SVD	SB	0.27	0.43

For each protein we compared the SVD and SB ensembles, the first of which was determined using RDC restraints calculated with the SVD method, and the second using the SB method (column 2). For each of these ensembles, we back-calculated the RDCs using the SVD method or the SB method (column 1). For example, the ensemble of ubiquitin determined using SVD-calculated RDCs has a $\bar{Q}$ factor of 0.61 when the RDCs are back-calculated using the SB method (row 2). The values reported in Table 2 differ from those in Table 1 because in the case of Q factors, the averaging over an ensemble of conformations is done for the RDC values, whereas for the $\bar{Q}$ factors the averaging over the ensemble is done directly on the Q factors

$$Q = \frac{\sqrt{\sum_i \left(D_i^{exp} - D_i^{calc}\right)^2}}{\sqrt{\sum_i \left(D_i^{exp}\right)^2}} \tag{14}$$

where the sums are taken over the number N_{RDCs} of available RDCs. In the case of ensembles of structures the Q factor is calculated by considering the average RDC values over the ensemble. The Q factor of the SVD ensemble (0.163) does not essentially improve with respect to that of the 1D3Z structure (0.164), which is not aimed at representing the dynamics, but rather the average structure of the protein. For comparison, the Q factor of a recently reported ensemble of ubiquitin determined also with a SVD approach (2KOX) (Fenwick et al. 2011) is 0.186, and the Q factor of another ensembles of ubiquitin recently reported (2K39) (Lange et al. 2008) is 0.216. All the Q factors reported in Fig. 2b and mentioned in this section were calculated as an average over the 50 alignment media that we considered in this work (Table S1). The ensemble of structures determined with the

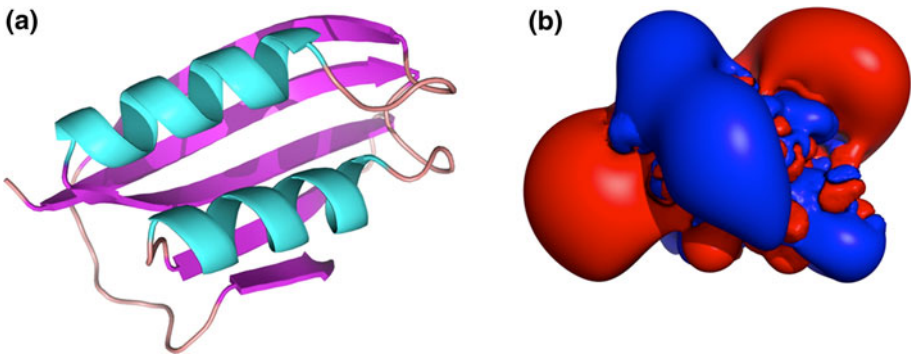


Fig. 4 **a** X-ray structure of Sso AcP (2BJD) (Corazza et al. 2006). **b** Representation of the electrostatic field on the surface of Sso AcP

method described in this work capture the dynamics of ubiquitin better than other methods, with a Q factor of 0.129. The use of only 4 sets of RDCs within the SB approach results in a slight loss of accuracy, but it still provides Q factors lower than those obtained using the SVD approach a much larger number (12 sets) of RDCs (Table 1).

Comparison of alternative methods for calculating alignment tensors

In order further investigate the reasons for which SVD methods are less effective in recovering the dynamics than the SB method that we describe here, we calculated the Q factors for the conformations in the SVD and SB ensembles either by using the SVD method or the SB method (Fig. 3; Table 2). We found that for both the SVD (Fig. 3a) and SB (Fig. 3b) ensembles the Q factors are lower when the alignment tensor is calculated using the SVD method (see also Table 2). These findings can be understood by recalling that the SVD method does not provide the actual alignment tensor of a given structure, but rather the alignment tensor that generates the RDC values in closest agreement with the experimental ones. In this sense the SB methods will never generate lower Q factors than SVD methods for individual structures, since the latter may suffer from overfitting problems. The extent of overfitting can be measured by considering the difference between the average Q factors, defined as

$$\Delta\bar{Q} = \bar{Q}_{SVD} - \bar{Q}_{SB} = \frac{1}{N_{struct}}\sum_{k=1}^{N_{struct}} Q_{k,SVD} - \frac{1}{N_{struct}}\sum_{k=1}^{N_{struct}} Q_{k,SB} \quad (15)$$

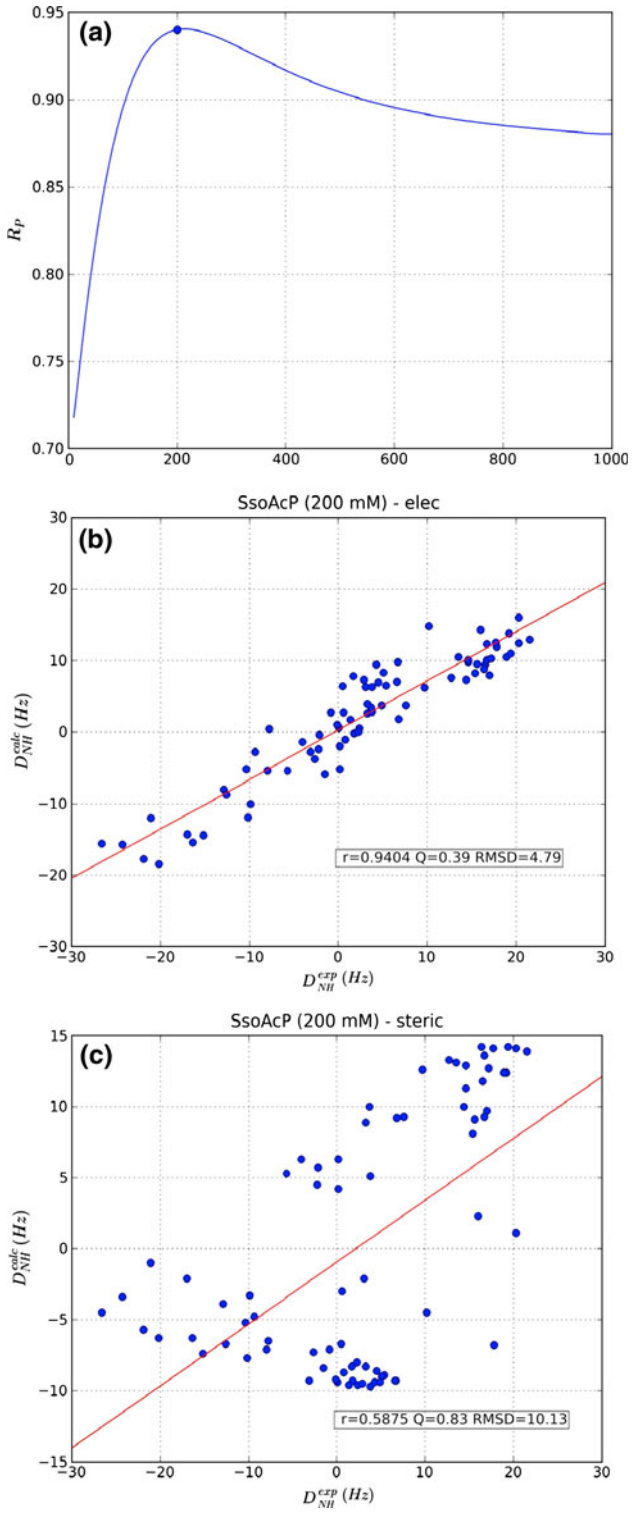
where $\bar{Q}_{SVD}$and $\bar{Q}_{SB}$ are the average Q factors obtained by using the SVD and the SB back-calculations, respectively; $\bar{Q}_{k,SVD}$and $\bar{Q}_{k,SB}$ are the Q factors for conformation k in an ensemble. The back-calculation in Table 2 is performed on individual structures of the ensemble both for the SB and single value decomposition (SVD) estimation of the alignment tensor. The resulting RDC values are then averaged over the ensemble. The results that we obtained indicate that the SVD ensemble appears to be of good quality in terms of Q factors when the back-calculation is performed with the SVD method, but of rather poor quality when the back-calculation is performed with the SB method. For the SB method this difference is $\Delta\bar{Q} = 0.17$ (Table 2, column 3, rows 4 and 5), and for the SVD method is $\Delta\bar{Q} = 0.27$ (Table 2, column 3, rows 2 and 3).

Determination of the structural fluctuations in the native state of Sso AcP

A particularly challenging case for SB methods is that of Sso AcP (Fig. 4a), since this protein has a very polarised charge distribution (Fig. 4b). The alignment of this protein in Pf1, which is the alignment media that we used in the RDC measurements (Table S2), is therefore strongly dependent on the ionic strength used in the experiment (Fig. 5a–c). By varying the ionic strength (NaCl) used in the SB calculation of the alignment tensor (Fig. 4a) we obtain the best agreement at 200 mM (Q factor 0.39, Fig. 5b), which is indeed the ionic strength that we used in the experiments. By increasing the ionic strength value in the calculations we obtain RDC values that converge to those corresponding to a steric medium (Fig. 5a). The importance of the electrostatic interactions is demonstrated by the calculations of the RDCs in Pf1 with the electrostatic terms switched off, which result in a Q factor of 0.83 (Fig. 5c).

The SB calculations shown in Fig. 5 were carried out using the X-ray structure of Sso AcP (2BJD) (Corazza et al. 2006), and hence they do not take account of the structural fluctuations of this protein. In order to include this aspect, we performed molecular dynamics simulations with replica-averaged SB RDC restraints, using the same procedure described above for ubiquitin, with the difference that in this case we used Pf1 in the SB calculations. These simulations provide an ensemble of conformations (Fig. 6a) whose Q factor demonstrates an excellent agreement between calculated and experimental RDCs (coefficient of correlation 0.9955, Q factor 0.11, Fig. 5b).

An analysis of the results presented in Table 2 indicates that the problem of overfitting introduced by the use of SVD methods to calculate restraints in molecular dynamics



◀ **Fig. 5** **a** Correlation between experimental (measured in Pf1 at 200 mM ionic strength) and structure-based RDCs; the latter were calculated from the X-ray structure of Sso AcP for values of the ionic strength ranging from 0 to 1,000 mM. The best correlation is found for 200 mM ionic strength, which matches the value used in the experiments. **b** *Scatter plot* of the correlation between experimental and structure-based RDCs at 200 mM ionic strength. **c** *Scatter plot* of the correlation between experimental and structure-based RDCs at 200 mM ionic strength, but with the electrostatic interactions switched off in the calculations

simulation is more severe for Sso AcP than for ubiquitin. In this case the best average Q factor is found for the SVD ensemble when the back-calculation is done with the SVD method (Table 2, column 4, row 3). However, there is a very wide difference ($\Delta\bar{Q} = 0.34$) between the Q factors back-calculated with the SB and SVD methods. By contrast, for the SB ensemble there is a much smaller corresponding difference ($\Delta\bar{Q} = 0.12$), indicating that the implementation of RDCs as structural restraints suffers less from overfitting. These results are consistent with the observation made above that the SVD method provides the alignment tensor that corresponds to the RDC values in closest agreement with the experimental ones, instead of the actual alignment tensor of a given structure.

As in the case of ubiquitin, therefore, by extracting the information about the conformational fluctuations provided by RDC data, we have been able to determine an ensemble of conformation that represents the native structural fluctuations of Sso AcP.

Discussion

In this work we have described a strategy aimed at extracting the information about conformational fluctuations of proteins provided by RDC measurements by incorporating them as structural restraints in molecular dynamics simulations. The method is based on the definition of an accurate mapping between RDCs and protein structures, which enables one to go from protein structures to RDCs and vice versa. In the following we discuss some important aspects of the calculations that we presented.

Assumptions made in the molecular dynamics simulations with RDC restraints

In this work we did not make the assumption that the alignment tensor of a protein is nearly constant during the dynamics (Clore and Schwieters 2004a, b; De Simone et al. 2009; Fenwick et al. 2011), nor that its fluctuations are uncorrelated with those of the protein (Salvatella et al. 2008). There are, however, other assumptions that we made to determine the dynamics of proteins using RDC restraints in molecular dynamics simulations:

1. *Only the orientation of a protein, but not its conformational properties, is changed by its interaction with the alignment medium.* This approximation can be expected to hold if the inter-molecular interactions between the protein and the alignment medium are weaker than the intra-molecular interactions within the protein. In this

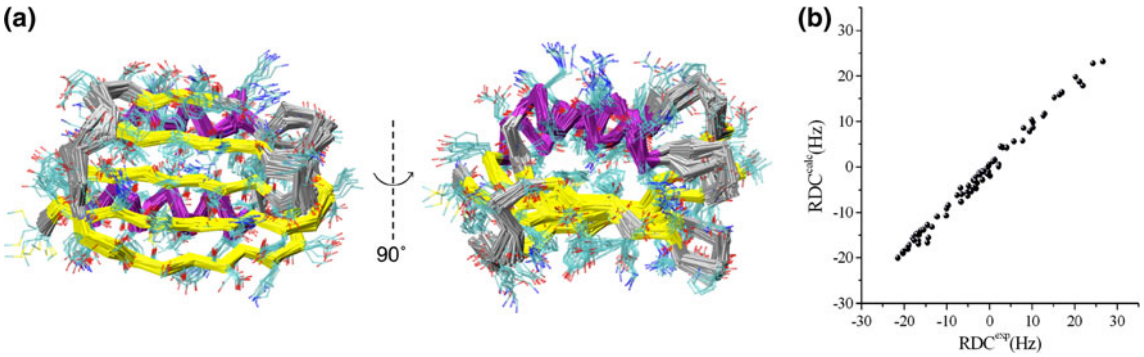


Fig. 6 **a** Ensemble of structures of Sso AcP calculated from N-HN RDCs measured in Pf1 and used as restraints in the approach presented in this work. **b** *Scatter plot* of the correlation between experimental and structure-based RDCs at 200 mM ionic strength for the ensemble of conformations shown in **a**; the coefficient of correlation is 0.9955

case the interactions with the alignment medium will affect much more strongly the orientation of the protein than its internal structure and dynamics. This assumption may be problematic in some cases, in particular for example for unstructured proteins. If this approximation cannot be made, it means that the process by with RDCs are measured perturbs significantly the conformational properties of a protein and therefore less invasive structural measurements may be considered.

2. *The partial charges on the protein do not change during the dynamics.* Although is well established that the partial charges of a protein are conformation dependent, we have assumed that this effect can be neglected in the cases of ubiquitin and Sso AcP discussed in this work. If needed, however, the procedure presented here can be implemented with a recalculation of the partial charges during the dynamics, in order to avoid this assumption.
3. *The electrostatic interactions with the alignment medium do not change the partial charges on the protein.* This approximation is valid if the electrostatic interaction between the protein and the alignment medium is weak, a situation that can be achieved by varying the ionic strength. In practice this means that only a specific range of values of the ionic strength are allowed for Pf1 (between 100 and 300 mM). In this case the Gouy-Chapman approximation holds to calculate the electrostatic field. It this approximation breaks down, one can expect strong interactions between the protein and the alignment medium, which would not be desirable as they would affect the conformational properties of the protein itself.

Structure-based (SB) calculations of the alignment tensor

The alignment tensor of a given protein conformation can be obtained through fitting procedures, such as the singular-value decomposition (SVD) method (Losonczi et al. 1999), in which the alignment tensor is chosen to optimise the agreement between calculated and experimental RDCs. Alternatively the alignment tensor can be determined by SB procedures in which this quantity is calculated on the basis of the shape and charge of the protein molecule (Zweckstetter and Bax 2000; Fernandes et al. 2001; Almond and Axelsen 2002; Azurmendi and Bush 2002; van Lune et al. 2002; Ferrarini 2003; Zweckstetter 2008; Berlin et al. 2009), without reference to experimentally measured RDCs.

These two general approaches are not equivalent in the presence of conformational fluctuations of large amplitude. In order to clarify this point, we observe that the calculation of the average RDCs corresponding to an ensemble of conformations involves the definition of the alignment tensor for each conformation in the ensemble (see Eq. 2). In approaches in which the RDCs are fitted to a structure, one can either assume that all the conformations in the ensemble have the same alignment tensor, or that the alignment tensor of each individual conformation can be obtained by a separate fitting to the experimental RDCs. In this latter case, a very large number of experimental RDCs is required in order to avoid overfitting. In any case, fitting methods are at risk of failing to capture the full changes in the alignment tensor during the conformational fluctuations (Fig. 3; Table 1). By contrast, the SB approach that we adopted in this work defines a specific alignment tensor for each conformation in the ensemble without relying on the knowledge of experimental RDCs.

Conclusions

We have presented a method of determining the structural fluctuations of proteins using information derived from RDC measurements. A key question that can be asked about any method of this type is whether it enables one to determine ensembles of structures with Q factors lower

than those that can be obtained from high-resolution individual structures, in particular for RDC data that are not used as restraints in the calculations. The results that we presented indicate that our method is capable of achieving this result.

Acknowledgments We are grateful to Francesco Aprile, Fabrizio Chiti and Giuliana Fusco for providing us with the Sso AcP samples for the RDC measurements and for assistance in the experiments. This work was supported by the BBSRC (Rinaldo W. Montalvao and Michele Vendruscolo), the EPSRC (Alfonso De Simone), the European Molecular Biology Organisation, and the Royal Society (Michele Vendruscolo).

References

Allison JR, Varnai P, Dobson CM, Vendruscolo M (2009) Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. J Am Chem Soc 131:18314–18326

Almond A, Axelsen JB (2002) Physical interpretation of residual dipolar couplings in neutral aligned media. J Am Chem Soc 124: 9986–9987

Azurmendi HF, Bush CA (2002) Tracking alignment from the moment of inertia tensor (tramite) of biomolecules in neutral dilute liquid crystal solutions. J Am Chem Soc 124:2426–2427

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

Berlin K, O'Leary DP, Fushman D (2009) Improvement and analysis of computational methods for prediction of residual dipolar couplings. J Magn Res 201:25–33

Bertoncini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, Zweckstetter M (2005) Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. Proc Natl Acad Sci USA 102:1430–1435

Best RB, Vendruscolo M (2004) Determination of protein structures consistent with NMR order parameters. J Am Chem Soc 126: 8090–8091

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

Boehr DD, McElheny D, Dyson HJ, Wright PE (2006) The dynamic energy landscape of dihydrofolate reductase catalysis. Science 313:1638–1642

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

Clore GM, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem Rev 109:4108–4139

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

Clore GM, Schwieters CD (2004b) 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

Corazza A, Rosano C, Pagano K, Alverdi V, Esposito G, Capanni C, Bemporad F, Plakoutsi G, Stefani M, Chiti F, Zuccotti S, Bolognesi M, Viglino P (2006) Structure, conformational stability, and enzymatic properties of acylphosphatase from the hyperthermophile *Sulfolobus solfataricus*. Proteins 62:64–79

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 Simone A, Richter B, Salvatella X, Vendruscolo M (2009) Toward an accurate determination of free energy landscapes in solution states of proteins. J Am Chem Soc 131:3810–3811

Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM (2005) Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J Am Chem Soc 127:476–477

Fenwick RB, Esteban-Martin S, Richter B, Lee D, Walter KFA, Milovanovic D, Becker S, Lakomek NA, Griesinger C, Salvatella X (2011) Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. J Am Chem Soc 133:10336–10339

Fernandes MX, Bernado P, Pons M, de la Torre JG (2001) An analytical solution to the problem of the orientation of rigid particles by planar obstacles. Application to membrane systems and to the calculation of dipolar couplings in protein NMR spectroscopy. J Am Chem Soc 123:12037–12047

Ferrarini A (2003) Modeling of macromolecular alignment in nematic virus suspensions. Application to the prediction of NMR residual dipolar couplings. J Phys Chem B 107:7923–7931

Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254:1598–1603

Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447

Hildebrandt A, Dehof AK, Rurainski A, Bertsch A, Schumann M, Toussaint NC, Moll A, Stockel D, Nickels S, Mueller SC, Lenhof HP, Kohlbacher O (2010) BALL—biochemical algorithms library 1.3. BMC Bioinform 11:531

Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins 65: 712–725

Huang JR, Grzesiek S (2010) Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin. J Am Chem Soc 132:694–705

Iwahara J, Schwieters CD, Clore GM (2004) Ensemble approach for NMR structure refinement against H1 paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. J Am Chem Soc 126:5879–5896

Iwahara J, Zweckstetter M, Clore GM (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. Proc Natl Acad Sci USA 103: 15062–15067

Kalodimos CG (2011) NMR reveals novel mechanisms of protein activity regulation. Protein Sci 20:773–782

Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KFA, Becker S, Meiler J, Grubmuller H, Griesinger C, de Groot BL (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 320: 1471–1475

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

Losonczi JA, Andrec M, Fischer MWF, Prestegard JH (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. J Magn Res 138:334–342

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, 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

Mittermaier A, Kay LE (2006) New tools provide new insights in NMR studies of protein dynamics. Science 312:224–228

Palmer AG (2004) NMR characterization of the dynamics of biomacromolecules. Chem Rev 104:3623–3640

Richter B, Gsponer J, Varnai 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

Salmon L, Bouvignies G, Markwick P, Blackledge M (2011) Nuclear magnetic resonance provides a quantitative description of protein conformational flexibility on physiologically important time scales. Biochemistry 50:2735–2747

Salvatella X, Richter B, Vendruscolo M (2008) Influence of the fluctuations of the alignment tensor on the analysis of the structure and dynamics of proteins using residual dipolar couplings. J Biomol NMR 40:71–81

Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan YB, Wriggers W (2010) Atomic-level characterization of the structural dynamics of proteins. Science 330:341–346

Showalter SA, Bruschweiler R (2007) Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. J Am Chem Soc 129:4158–4159

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 JR, 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

van Lune F, Manning L, Dijkstra K, Berendsen HJC, Scheek RM (2002) Order-parameter tensor description of HPr in a medium of oriented bicelles. J Biomol NMR 23:169–179

Vendruscolo M (2007) Determination of conformationally heterogeneous states of proteins. Curr Opin Struct Biol 17:15–20

Vendruscolo M, Dobson CM (2006) Dynamic visions of enzymatic reactions. Science 313:1586–1587

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

Zhang FL, Bruschweiler R (2002) Contact model for the prediction of NMR NH order parameters in globular proteins. J Am Chem Soc 124:12654–12655

Zhang Q, Sun XY, Watt ED, Al-Hashimi HM (2006) Resolving the motional modes that code for RNA adaptation. Science 311: 653–656

Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. Nature 450:U1214–U1263

Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. Nat Prot 3:679–690

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, Hummer G, Bax A (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. Biophys J 86:3444–3460