

ORIUM: Optimized RDC-based Iterative and Unified Model-free analysis

T. Michael Sabo · Colin A. Smith · David Ban ·
Adam Mazur · Donghan Lee · Christian Griesinger

Received: 22 May 2013 / Accepted: 23 August 2013 / Published online: 8 September 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract Residual dipolar couplings (RDCs) are NMR parameters that provide both structural and dynamic information concerning inter-nuclear vectors, such as N–H^N and C α –H α bonds within the protein backbone. Two approaches for extracting this information from RDCs are the model free analysis (MFA) (Meiler et al. in *J Am Chem Soc* 123:6098–6107, 2001; Peti et al. in *J Am Chem Soc* 124:5822–5833, 2002) and the direct interpretation of dipolar couplings (DIDCs) (Tolman in *J Am Chem Soc* 124:12020–12030, 2002). Both methods have been incorporated into iterative schemes, namely the self-consistent RDC based MFA (SCRM) (Lakomek et al. in *J Biomol NMR* 41:139–155, 2008) and iterative DIDC (Yao et al. in *J Phys Chem B* 112:6045–6056, 2008), with the goal of removing the influence of structural noise in the MFA and DIDC formulations. Here, we report a new iterative procedure entitled Optimized RDC-based Iterative and Unified

Model-free analysis (ORIUM). ORIUM unifies theoretical concepts developed in the MFA, SCRM, and DIDC methods to construct a computationally less demanding approach to determine these structural and dynamic parameters. In all schemes, dynamic averaging reduces the actual magnitude of the alignment tensors complicating the determination of the absolute values for the generalized order parameters. To readdress this scaling issue that has been previously investigated (Lakomek et al. in *J Biomol NMR* 41:139–155, 2008; Salmon et al. in *Angew Chem Int Edit* 48:4154–4157, 2009), a new method is presented using only RDC data to establish a lower bound on protein motion, bypassing the requirement of Lipari–Szabo order parameters. ORIUM and the new scaling procedure are applied to the proteins ubiquitin and the third immunoglobulin domain of protein G (GB3). Our results indicate good agreement with the SCRM and iterative DIDC approaches and signify the general applicability of ORIUM and the proposed scaling for the extraction of inter-nuclear vector structural and dynamic content.

T. Michael Sabo and Colin A. Smith have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-013-9775-1) contains supplementary material, which is available to authorized users.

T. M. Sabo · D. Ban · A. Mazur · D. Lee (✉) ·
C. Griesinger (✉)

Department for NMR-based Structural Biology, Max-Planck
Institute for Biophysical Chemistry, Am Fassberg 11,
37077 Göttingen, Germany
e-mail: dole@nmr.mpibpc.mpg.de

C. Griesinger
e-mail: cigr@nmr.mpibpc.mpg.de

C. A. Smith
Department of Theoretical and Computational Biophysics, Max-
Planck Institute for Biophysical Chemistry, Am Fassberg 11,
37077 Göttingen, Germany

Keywords RDCs · Dynamics · Ubiquitin · GB3

Introduction

Protein structure and dynamics are routinely investigated with NMR spectroscopy at atomic resolution. An essential NMR parameter that provides both structural and dynamic information is the residual dipolar coupling (RDC) between two nuclear magnetic moments, for example a N–H^N or C α –H α vector within the polypeptide backbone of a protein (Tjandra and Bax 1997; Tolman et al. 1997). An important application of RDCs as a probe for protein dynamics has been shown recently where RDCs measured

in 36 different alignment media demonstrated that the ground state conformational ensemble of ubiquitin covers the conformational space captured in crystal structures of ubiquitin complexes (Lange et al. 2008; Lakomek et al. 2008). These findings provide strong support for conformational selection as a means for molecular recognition. Therefore, extracting the dynamical content from RDCs has implications for understanding the mechanisms of molecular recognition and protein function.

The significance of the RDC's dynamical content is highlighted when considering other approaches for studying protein dynamics. Measurements of NMR spin-relaxation provide information concerning amplitudes of inter-nuclear vector motions occurring on time-scales faster than the rotational correlation time (τ_c) of the protein (picosecond to nanosecond) (Kay et al. 1989), which are parameterized by the Lipari–Szabo order parameter (S_{LS}^2) (Lipari and Szabo 1982a, b). Relaxation dispersion methods probe the kinetics of conformational exchange that modulates the isotropic chemical shift and contributes to the effective transverse relaxation rate (Palmer 2004). To date, relaxation dispersion techniques have been limited to time-scales slower than approximately 25 μ s (Ban et al. 2011, 2012). By contrast, RDCs provide vital insight into the amplitude and direction of internal vector motions on time-scales covering the previously inaccessible time window spanning τ_c to ~ 25 μ s (referred to as the supra- τ_c window).

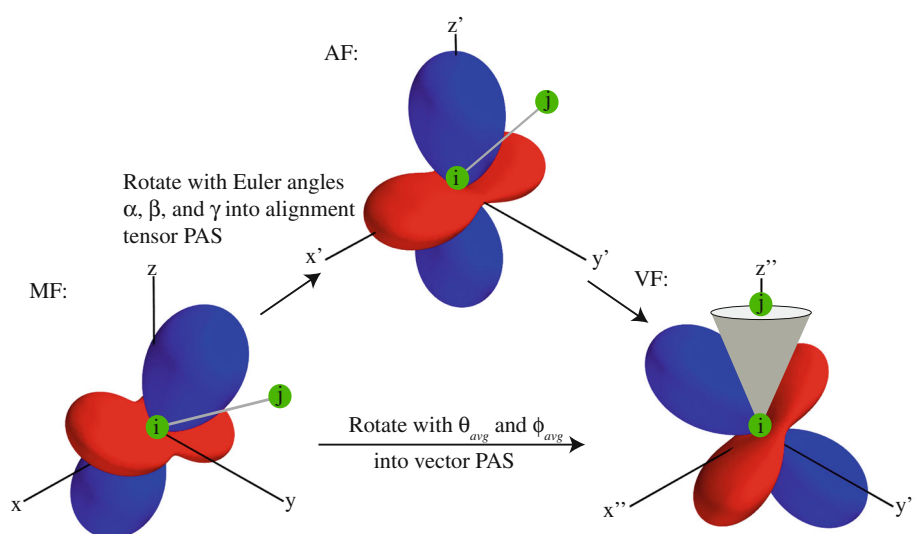
Residual dipolar couplings (RDCs) arise from placing a protein in an anisotropic medium, such as filamentous phages or lipid bilayers, or paramagnetic tagging, leading to partial alignment of the protein with respect to the external magnetic field. In the anisotropic media, all possible orientations for an inter-nuclear vector are populated with unequal probability, resulting in the dipolar couplings no longer averaging to zero. The magnitude of the measured RDC is given by the time-averaged angle between

the inter-nuclear vector and the magnetic field (Tolman et al. 1997).

Since the potential to extract dynamics from RDCs was first recognized, two schemes for extracting the dynamical content from these NMR parameters in the form of a generalized order parameter (S_{RDC}^2) have been proposed. In the model free analysis (MFA), five independent alignment media are necessary to calculate the five independent elements of the inter-nuclear vector tensor (Meiler et al. 2001; Peti et al. 2002). Figure 1 illustrates the three frames of reference used in the analysis of RDCs, the molecular frame (MF), the alignment frame (AF), and the vector frame (VF). Knowledge of the protein structure is necessary to determine the alignment tensors. With the alignment tensor information, the averages over the second rank spherical harmonics describing the mean orientations of the vectors, contained within the inter-nuclear vector tensor ($\langle Y_{2,m}(\theta, \phi) \rangle$), see also Fig. 2), provide the desired structural and dynamic content. An alternative approach, the direct interpretation of dipolar couplings (DIDCs), was developed to bypass the need for structural input in the calculation of the inter-nuclear vector's structural and dynamic content (Tolman 2002). Five independent alignment media are also necessary for the DIDC method. A single matrix equation is employed to represent the RDC data obtained in multiple alignment media. The inter-nuclear vector tensors are optimized simultaneously and variation in S_{RDC}^2 is minimized.

Both the MFA and DIDC methods have been incorporated into iterative schemes with the goal of improving the accuracy of the alignment tensor calculation by reducing the effects of the structural noise, termed the self-consistent RDC based MFA (SCRM) and iterative DIDC (Lakomek et al. 2008; Yao et al. 2008). The iterative schemes achieve this by using the refined dynamically averaged coordinates as input for additional runs of either MFA or DIDC, however

Fig. 1 Illustration of the molecular frame (MF), the alignment frame (AF), and the vector frame (VF) and the relationship between the different frames of reference. Conceptually, the SCRM procedure moves from the MF to the AF to the VF, while ORIUM goes directly from the MF to the VF. The alignment tensor is depicted with the positive lobes in blue and the negative lobes in red



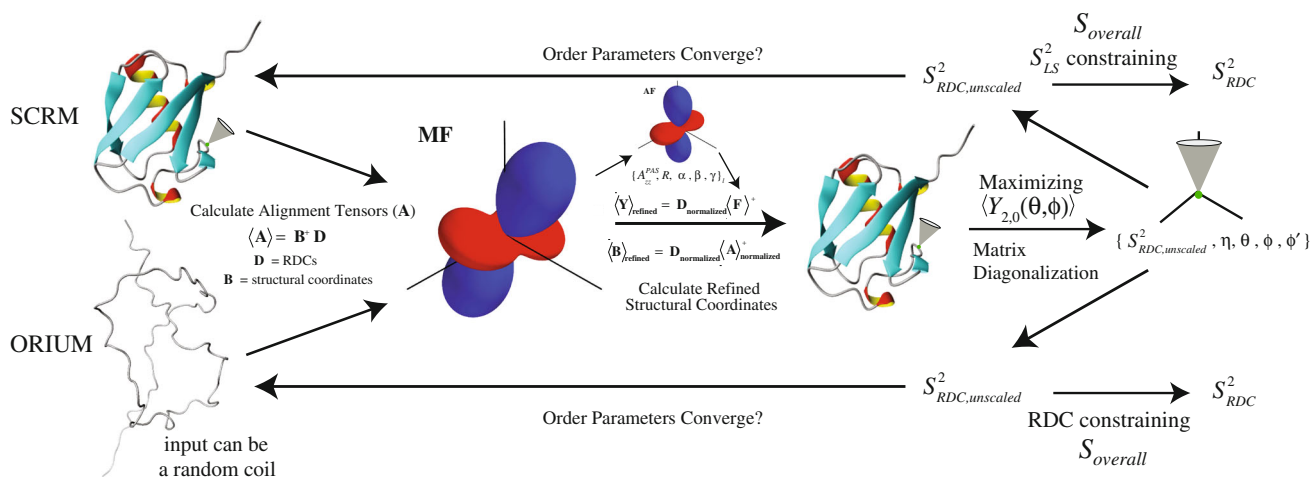


Fig. 2 Schematic diagram of both ORIUM and SCRM protocols (Lakomek et al. 2008). Both procedures begin with a starting structure as input and use experimental RDCs to calculate the alignment tensors in the molecular frame (MF). In SCRM, the principal axis system of each alignment tensor is determined in order to extract $\{A_{zz}^{PAS}, R, \alpha, \beta, \gamma\}_l$ which are used to construct the **F** matrix and scale the RDCs. With ORIUM, only A_{zz}^{PAS} is needed to scale the RDCs. The next step is to calculate refined structural coordinates. SCRM requires the five alignment tensor parameters $\{A_{zz}^{PAS}, R, \alpha, \beta, \gamma\}_l$ to construct

the **F** matrix for the determination of $\langle Y \rangle_{\text{refined}}$. ORIUM determines $\langle B \rangle_{\text{refined}}$ directly from $\langle A \rangle_{\text{normalized}}$. In order to calculate $\{S^2_{RDC}, \eta, \theta^{MF}, \phi^{MF}, \phi'\}_k$, SCRM maximizes $\langle Y_{2,0}(\theta_k^{VF}, \phi_k^{VF}) \rangle$ while ORIUM utilizes Sauepe matrix diagonalization (Meirovitch et al. 2012). The angles $(\theta_k^{MF}, \phi_k^{MF})$ are then used as input to restart the SCRM and ORIUM protocols until $S^2_{RDC,unscaled}$ converges. Once this criterion is fulfilled, the $S^2_{overall}$ parameter is determined, which scales the order parameters

each approach, as implemented, relies on computationally expensive procedures. In the iterative DIDC method, a grid search is performed which minimizes the difference between the vector’s coordinates and a pool of possible solutions built from an exhaustive list of (θ, ϕ) combinations to find dynamically averaged coordinates (Yao et al. 2008). As for SCRM, the dynamic average orientation of each vector is calculated by performing a coordinate transformation with maximization of $\langle Y_{2,0}(\theta, \phi) \rangle$ (Lakomek et al. 2008). Recently, it has been shown that this transformation can be replaced with the diagonalization of the local ordering Sauepe matrix (Meirovitch et al. 2012), which is computationally less demanding.

Here, we describe a new iterative procedure for extracting structural and dynamic information from RDCs entitled Optimized RDC-based Iterative and Unified Model-free analysis (ORIUM). ORIUM unifies the theoretical concepts developed in the MFA, SCRM, and DIDC methods. In addition, a new method is presented to establish a lower bound on protein motion using RDC data alone without requiring a separate determination of S^2_{LS} . While this has been achieved previously (Salmon et al. 2009) based on several sets of RDCs assuming Gaussian fluctuations, the method introduced here works on a single set of RDCs and does not require a motional model. The applicability of ORIUM and the new scaling procedure are tested with the model proteins ubiquitin and the third immunoglobulin domain of protein G (GB3).

Theory

Optimized RDC-based Iterative and Unified Model-free analysis (ORIUM) consists of three principal stages for the extraction of RDC order parameters (S^2_{RDC}) from data measured in multiple alignment media (see Fig. 2 for schematic diagram). First, the matrix formalism introduced by Tolman in the DIDC approach is utilized to calculate refined structural coordinates from the alignment tensors (Tolman 2002). From here, each refined vector is put into a local axis system in order to determine the vector specific structural and dynamic information (Meirovitch et al. 2012; Meiler et al. 2001; Peti et al. 2002). Finally, the resulting Euler angles are used as structural input to restart the calculation in an iterative fashion, similar to SCRM (Lakomek et al. 2008). ORIUM continues until the variation in (S^2_{RDC}) for the entire dataset falls below a certain threshold.

Alignment tensor calculation

For two nuclear spins, the observed resonance splitting (Hz) resulting from the partial alignment of a protein emanate from the secular part of the magnetic dipole interaction

$$D_k^{exp} = D_{ij}^{max} \langle (3 \cos^2 \theta_k - 1) / 2 \rangle \tag{1}$$

$$D_{ij}^{max} = -\frac{\mu_0 \gamma_i \gamma_j \hbar}{4\pi^2 r_{ij}^3} \tag{1a}$$

where μ_0 is the permeability of vacuum, γ_X is the gyro-magnetic ratio of spin X , \hbar is Planck's constant, r_{ij} is the distance between nuclei i and j (assumed to be fixed at 1.02 Å for the N–H^N and 1.095 Å for the C α –H α vectors), and θ_k is the angle between the inter-nuclear vector formed by nuclear spin pair k and the magnetic field (B_0). The angular brackets denote ensemble averaging. As Eq. (1) explicitly illustrates, the magnitude of D_k^{exp} depends on $\langle(3 \cos^2 \theta_k - 1)/2\rangle$. By definition, the term $\cos \theta_k$ is the scalar product between an inter-nuclear vector and the vector parallel to B_0 .

When considering a rigid molecule, the coordinates of an inter-nuclear vector can be described within an arbitrary reference frame, termed the molecular frame (MF), and defined by three angles, β_x , β_y , and β_z , between the vector and the respective MF axes. In a similar fashion, the vector parallel to B_0 can be expressed by three angles representing the instantaneous orientation of B_0 relative to the MF axes, α_x , α_y , and α_z . Within the MF, D_k^{exp} can be recast as

$$D_k^{exp} = D_{ij}^{max} \langle B \cdot A \rangle \quad (2)$$

where $\langle B \cdot A \rangle$ is the scalar product of two vectors representing the inter-nuclear orientations (B) and the B_0 orientations (A). Here, A is the alignment tensor and B is the inter-nuclear vector tensor. Both A and B contain five independent terms and are related to a 3×3 second rank Cartesian order tensor as follows (Saupe 1964, 1968; Snyder 1965)

$$A = \left[a_{zz}, \frac{1}{\sqrt{3}}(a_{xx} - a_{yy}), \frac{2}{\sqrt{3}}a_{xz}, \frac{2}{\sqrt{3}}a_{yz}, \frac{2}{\sqrt{3}}a_{xy} \right] \quad (3)$$

where the orientation of B_0 in the MF is given by

$$a_{mn} = \left\langle \frac{1}{2}(3 \cos \alpha_m \cos \alpha_n - \delta_{mn}) \right\rangle_l \quad (3a)$$

and

$$B = \left[b_{zz}^k, \frac{1}{\sqrt{3}}(b_{xx}^k - b_{yy}^k), \frac{2}{\sqrt{3}}b_{xz}^k, \frac{2}{\sqrt{3}}b_{yz}^k, \frac{2}{\sqrt{3}}b_{xy}^k \right] \quad (4)$$

where the orientation of the inter-nuclear vector in the MF is described by

$$b_{mn}^k = \left\langle \frac{1}{2}(3 \cos \beta_m^k \cos \beta_n^k - \delta_{mn}) \right\rangle. \quad (4a)$$

The term δ_{mn} represents the Kronecker delta function, l is the alignment condition, and $m, n = x, y, z$.

A matrix formalism is introduced to render analysis of the RDC data in a more intuitive manner (Tolman 2002). When K RDCs are measured under L alignments, then Eq. (2) becomes

$$\mathbf{D} = \langle \mathbf{B} \rangle \langle \mathbf{A} \rangle \quad (5)$$

where \mathbf{D} is a $K \times L$ matrix, \mathbf{B} is a $K \times 5$ matrix, and \mathbf{A} is a $5 \times L$ matrix. In Eq. (5), the term D_{ij}^{max} is included in $\langle \mathbf{A} \rangle$. The rows of \mathbf{B} are defined by Eq. (4) and the columns of \mathbf{A}

are given by Eq. (3). An inherent assumption in the present analysis is that inter-nuclear dynamics are uncorrelated with the alignment process; hence the averages of $\langle \mathbf{B} \rangle$ and $\langle \mathbf{A} \rangle$ are independent of each other. This assumption can be tested with the SECONDA analysis (Hus and Brüschweiler 2002; Hus et al. 2003). When the structure of the molecule is known and RDCs for at least five linearly independent inter-nuclear vectors are measured, the matrix \mathbf{B} (input from the rigid structure or random structural coordinates) and the measured RDCs are used to calculate $\langle \mathbf{A} \rangle$

$$\langle \mathbf{A} \rangle = \mathbf{B}^+ \mathbf{D} \quad (6)$$

where \mathbf{B}^+ is the pseudo-inverse of \mathbf{B} . It should be noted that a single alignment tensor per alignment medium is necessary for the successful application of the following protocols. Intrinsically disordered proteins (see Bertocini et al. 2005; Bernadó et al. 2005) and multiple domain proteins (see Bertini et al. 2004; Rodriguez-Castañeda et al. 2006) will have to be described by several alignment tensors per alignment medium and will not be amenable to the present analysis.

Each column of $\langle \mathbf{A} \rangle$, given by Eq. (3), can be recast into L symmetric 3×3 second rank Cartesian order tensors, ($A_l^{(2)}$). These order matrices are then redefined in a principal axis system (PAS), termed the alignment frame (AF), where Eq. (1) becomes (Bax et al. 2001)

$$D_{k,l}^{exp} = D_{a,l} \left[\left\langle 3 \cos^2 \theta_{k,l}^{AF} - 1 \right\rangle + \frac{3}{2} R_l \left\langle \sin^2 \theta_{k,l}^{AF} \cos 2\phi_{k,l}^{AF} \right\rangle \right]. \quad (7)$$

In Eq. (7), the magnitude of the alignment tensor is $D_{a,l} = \frac{1}{2} D_{ij}^{max} * A_{zz,l}^{PAS}$, the rhombicity is $R_l = \frac{2(A_{xx,l}^{PAS} - A_{yy,l}^{PAS})}{3 * A_{zz,l}^{PAS}}$, ($\theta_{k,l}^{AF}$, $\phi_{k,l}^{AF}$) are the polar angles defining the inter-nuclear vector in the AF, and $A_{(mm,l)}^{PAS}$ are the eigenvalues resulting from the diagonalization of $A_l^{(2)}$. From the eigenvectors ($A_{mn,l}^{EV}$), the Euler angles describing the rotation of $A_l^{(2)}$ into the PAS are defined

$$\begin{aligned} \alpha_l &= \arctan \left[A_{xz,l}^{EV}, A_{yz,l}^{EV} \right], & \beta_l &= \arccos \left[A_{zz,l}^{EV} \right], \\ \gamma_l &= \arctan \left[-A_{zx,l}^{EV}, A_{zy,l}^{EV} \right]. \end{aligned} \quad (8)$$

Model free analysis

With the MFA (Meiler et al. 2001; Peti et al. 2002), the five parameters describing each alignment tensor in the PAS, $\{A_{zz}^{PAS}, R, \alpha, \beta, \gamma\}$, are used to construct the \mathbf{F} matrix which is needed to derive the five dynamically averaged second order spherical harmonics

$$\langle Y_{2,0}(\theta_k^{MF}, \phi_k^{MF}) \rangle = \sqrt{\frac{5}{16\pi}} \langle 3 \cos^2 \theta_k^{MF} - 1 \rangle \tag{9a}$$

$$\langle Y_{2,\pm 1}(\theta_k^{MF}, \phi_k^{MF}) \rangle = \mp \sqrt{\frac{15}{8\pi}} \langle e^{\pm i\phi_k^{MF}} \cos \theta_k^{MF} \sin \theta_k^{MF} \rangle \tag{9b}$$

$$\langle Y_{2,\pm 2}(\theta_k^{MF}, \phi_k^{MF}) \rangle = \sqrt{\frac{15}{32\pi}} \langle e^{\pm 2i\phi_k^{MF}} \sin^2 \theta_k^{MF} \rangle. \tag{9c}$$

Equation (7) can be recast in terms of dynamically averaged second order spherical harmonics

$$D_{k,l}^{exp} = A_{z,l}^{PAS} \sqrt{\frac{4\pi}{5}} \left[\langle Y_{2,0}(\theta_{k,l}^{AF}, \phi_{k,l}^{AF}) \rangle + \sqrt{\frac{3}{8}} R_l \left(\langle Y_{2,2}(\theta_{k,l}^{AF}, \phi_{k,l}^{AF}) \rangle + \langle Y_{2,-2}(\theta_{k,l}^{AF}, \phi_{k,l}^{AF}) \rangle \right) \right] \tag{10}$$

where

$$\langle Y_{2,0}(\theta_{k,l}^{AF}, \phi_{k,l}^{AF}) \rangle = \sqrt{\frac{5}{16\pi}} \langle 3 \cos^2 \theta_{k,l}^{AF} - 1 \rangle \tag{10a}$$

$$\begin{aligned} & \langle Y_{2,2}(\theta_{k,l}^{AF}, \phi_{k,l}^{AF}) \rangle + \langle Y_{2,-2}(\theta_{k,l}^{AF}, \phi_{k,l}^{AF}) \rangle \\ &= 2 \sqrt{\frac{15}{32\pi}} \langle \sin^2 \theta_{k,l}^{AF} \cos 2\phi_{k,l}^{AF} \rangle. \end{aligned} \tag{10b}$$

The **F** matrix relates the measured RDCs to the spherical harmonics defined in the MF by a Wigner rotation from the MF to the AF

$$\frac{D_{k,l}^{exp}}{A_{z,l}^{PAS}} = \sum_{M=-2}^2 F_{l,M} \langle Y_{2,M}(\theta_k^{MF}, \phi_k^{MF}) \rangle \tag{11}$$

with

$$F_{l,M} = \sqrt{\frac{4\pi}{5}} \left[D_{M0}^2(\alpha_l, \beta_l, \gamma_l) + \sqrt{\frac{3}{8}} R_l (D_{M2}^2(\alpha_l, \beta_l, \gamma_l) + D_{M-2}^2(\alpha_l, \beta_l, \gamma_l)) \right] = \sqrt{\frac{4\pi}{5}} \left[e^{-iM\alpha_l} d_{M0}^2(\beta_l) + \sqrt{\frac{3}{8}} R_l (e^{-iM\alpha_l} d_{M2}^2(\beta_l) e^{-i2\gamma_l} + e^{-iM\alpha_l} d_{M-2}^2(\beta_l) e^{i2\gamma_l}) \right]. \tag{12}$$

In analogy to the component definition from Eq. (5), **Y** is a $K \times 5$ matrix containing the dynamically averaged spherical harmonics in the MF and **F** is a $5 \times L$ matrix

containing the alignment tensor information. The $\langle \mathbf{Y} \rangle_{\text{refined}}$ matrix is determined in direct correspondence to Eq. (6)

$$\langle \mathbf{Y} \rangle_{\text{refined}} = \mathbf{D}_{\text{normalized}} \langle \mathbf{F} \rangle^+ \tag{13}$$

Here, $\mathbf{D}_{\text{normalized}}$ represents $\frac{D_{k,l}^{exp}}{A_{z,l}^{PAS}}$ in order to normalize the contributions of each alignment condition to the calculation of refined structural coordinates. Each row of $\langle \mathbf{Y} \rangle_{\text{refined}}$ is used to determine $S_{RDC,k}^2$

$$S_{RDC,k}^2 = \frac{4\pi}{5} \sum_{M=-2}^2 \langle Y_{2,M}(\theta_k^{MF}, \phi_k^{MF}) \rangle \langle Y_{2,M}^*(\theta_k^{MF}, \phi_k^{MF}) \rangle. \tag{14}$$

From the dynamically averaged spherical harmonics, the dynamically averaged orientations for each inter-nuclear vector, $(\theta_{\text{avg},k}^{MF}, \phi_{\text{avg},k}^{MF})$, can be obtained. Maximizing $\langle Y_{2,0}(\theta_k^{VF}, \phi_k^{VF}) \rangle$ places the z axis of the vector's axis system, termed the vector frame (VF), in the center of the inter-nuclear vector's orientational distribution,

$$\begin{aligned} & \max \langle Y_{2,0}(\theta_k^{VF}, \phi_k^{VF}) \rangle \\ &= \sum_{M=-2}^2 D_{M,0}(\phi_{\text{avg},k}^{MF}, \theta_{\text{avg},k}^{MF}, 0) \langle Y_{2,M}(\theta_k^{MF}, \phi_k^{MF}) \rangle \\ &= \sqrt{\frac{4\pi}{5}} \sum_{M=-2}^2 Y_{2,-M}(\theta_{\text{avg},k}^{MF}, \phi_{\text{avg},k}^{MF}) \langle Y_{2,M}(\theta_k^{MF}, \phi_k^{MF}) \rangle. \end{aligned} \tag{15}$$

The terms $\langle Y_{2,\pm 1}(\theta_k^{MF}, \phi_k^{MF}) \rangle$ vanish in the VF and $\langle Y_{2,\pm 2}(\theta_k^{MF}, \phi_k^{MF}) \rangle$ possesses information on the amplitude of anisotropy, η_k , and the orientation of anisotropic motions, ϕ'_k

$$\eta_k = \sqrt{\frac{\sum_{M=-2,2} \langle Y_{2,M}(\theta_k^{VF}, \phi_k^{VF}) \rangle \langle Y_{2,-M}(\theta_k^{VF}, \phi_k^{VF}) \rangle}{\sum_{M=-2}^2 \langle Y_{2,M}(\theta_k^{VF}, \phi_k^{VF}) \rangle \langle Y_{2,-M}(\theta_k^{VF}, \phi_k^{VF}) \rangle}} \tag{16}$$

$$\phi'_k = \frac{1}{2} \arctan \frac{\langle Y_{2,2}(\theta_k^{VF}, \phi_k^{VF}) \rangle - \langle Y_{2,-2}(\theta_k^{VF}, \phi_k^{VF}) \rangle}{i(\langle Y_{2,2}(\theta_k^{VF}, \phi_k^{VF}) \rangle + \langle Y_{2,-2}(\theta_k^{VF}, \phi_k^{VF}) \rangle)}. \tag{17}$$

It should be noted that $S_{RDC,k}^2$ is the same in any frame, thus

$$S_{RDC,k}^2 = \frac{4\pi}{5} \sum_{M=-2}^2 \langle Y_{2,M}(\theta_k^{VF}, \phi_k^{VF}) \rangle \langle Y_{2,M}^*(\theta_k^{VF}, \phi_k^{VF}) \rangle, \tag{18}$$

which is equivalent to Eq. (14).

Standard tensorial analysis

Recalling Eq. (4), the following relationships are established in order to construct $B_k^{(2)}$ (Snyder 1965)

$$b_{zz,k} = \sqrt{\frac{4\pi}{5}} \langle Y_{2,0}(\theta_k^{MF}, \phi_k^{MF}) \rangle \quad (19a)$$

$$\frac{1}{\sqrt{3}} (b_{xx,k} - b_{yy,k}) = \sqrt{\frac{1}{2}} \sqrt{\frac{4\pi}{5}} [\langle Y_{2,-2}(\theta_k^{MF}, \phi_k^{MF}) \rangle + \langle Y_{2,2}(\theta_k^{MF}, \phi_k^{MF}) \rangle] \quad (19b)$$

$$\frac{2}{\sqrt{3}} b_{xz,k} = \sqrt{\frac{1}{2}} \sqrt{\frac{4\pi}{5}} [\langle Y_{2,-1}(\theta_k^{MF}, \phi_k^{MF}) \rangle - \langle Y_{2,1}(\theta_k^{MF}, \phi_k^{MF}) \rangle] \quad (19c)$$

$$\frac{2}{\sqrt{3}} b_{yz,k} = i \sqrt{\frac{1}{2}} \sqrt{\frac{4\pi}{5}} [\langle Y_{2,-1}(\theta_k^{MF}, \phi_k^{MF}) \rangle + \langle Y_{2,1}(\theta_k^{MF}, \phi_k^{MF}) \rangle] \quad (19d)$$

$$\frac{2}{\sqrt{3}} b_{xy,k} = i \sqrt{\frac{1}{2}} \sqrt{\frac{4\pi}{5}} [\langle Y_{2,-2}(\theta_k^{MF}, \phi_k^{MF}) \rangle - \langle Y_{2,2}(\theta_k^{MF}, \phi_k^{MF}) \rangle]. \quad (19e)$$

The resulting eigenvalues ($B_{nm,k}^{PAS}$) contain the dynamic information for each vector ($S_{RDC,k}^2, \eta_k$), while the eigenvectors, ($B_{mn,k}^{EV}$), encompass the bond orientations ($\theta_k^{MF}, \phi_k^{MF}$) and the direction of the anisotropic local motion (ϕ_k'). The following equations detail how the dynamic parameters are calculated from $B_{nm,k}^{PAS}$. The Saupe order parameters are defined as

$$S_{0,k}^2 = B_{zz,k}^{PAS} = \sqrt{\frac{4\pi}{5}} \langle Y_{2,0}(\theta_k^{PAS}, \phi_k^{PAS}) \rangle \quad (20a)$$

$$S_{2,k}^2 = \sqrt{\frac{2}{3}} (B_{xx,k}^{PAS} - B_{yy,k}^{PAS}) \\ = \sqrt{\frac{4\pi}{5}} [\langle Y_{2,2}(\theta_k^{PAS}, \phi_k^{PAS}) \rangle + \langle Y_{2,-2}(\theta_k^{PAS}, \phi_k^{PAS}) \rangle]. \quad (20b)$$

$$S_{RDC,k}^2 = (B_{zz,k}^{PAS})^2 + \frac{1}{3} (B_{xx,k}^{PAS} - B_{yy,k}^{PAS})^2 \\ = (S_{0,k}^2)^2 + \frac{1}{2} (S_{2,k}^2)^2 \quad (21)$$

$$\eta_k^{PAS} = \sqrt{\frac{\frac{1}{3} (B_{xx,k}^{PAS} - B_{yy,k}^{PAS})^2}{S_{RDC,k}^2}} = \sqrt{\frac{\frac{1}{2} (S_{2,k}^2)^2}{S_{RDC,k}^2}}. \quad (22)$$

For each inter-nuclear vector, ($\theta_k^{MF}, \phi_k^{MF}$) and ϕ_k' are extracted from the transpose of the resulting $B_{nm,k}^{EV}$ matrix

$$\phi_k^{MF} = \arctan [B_{xz,k}^{EV}, B_{yz,k}^{EV}], \quad \theta_k^{MF} = \arccos [B_{zz,k}^{EV}], \quad (23) \\ \phi_k' = \arctan [-B_{zx,k}^{EV}, B_{zy,k}^{EV}].$$

Direct interpretation of dipolar couplings

With DIDC, once $\langle \mathbf{A} \rangle$ is determined from Eq. (6), $\langle \mathbf{A} \rangle$ is used to directly calculate a new set of dynamically averaged coordinates, $\mathbf{B}_{\text{refined}}$, without extracting each set of $\{A_{zz,l}^{PAS}, R, \alpha, \beta, \gamma\}_l$, according to

$$\langle \mathbf{B} \rangle_{\text{refined}} = \mathbf{D} \langle \mathbf{A} \rangle^+ + \mathbf{B} [1 - \langle \mathbf{A} \rangle \langle \mathbf{A} \rangle^+]. \quad (24)$$

This formula leaves the information for $\langle \mathbf{A} \rangle$ in the MF. It should be noted that the previous implementations of DIDC did not scale the RDCs by $A_{zz,l}^{PAS}$ as in the MFA (see Eq. 13), which is necessary to normalize the contributions of each alignment condition for the calculation of refined structural coordinates. Therefore, we have modified Eq. (24) as follows

$$\langle \mathbf{B} \rangle_{\text{refined}} = \mathbf{D}_{\text{normalized}} \langle \mathbf{A} \rangle_{\text{normalized}}^+ + \mathbf{B} [1 - \langle \mathbf{A} \rangle \langle \mathbf{A} \rangle^+], \quad (25)$$

where $\mathbf{D}_{\text{normalized}}$ and $\langle \mathbf{A} \rangle_{\text{normalized}}$ represent the RDCs and alignment tensors divided by $A_{zz,l}^{PAS}$.

As described by Tolman, the first term in Eqs. (24) and (25) encompasses the contribution of the measured RDCs to determining $\langle \mathbf{B} \rangle_{\text{refined}}$ (Tolman 2002). When the rank of $\langle \mathbf{A} \rangle$ is smaller than 5, then the second term accounts for the degeneracy in the possible solutions that results from B. Otherwise, this term will equal zero for data representing more than five alignment media. With $\langle \mathbf{B} \rangle_{\text{refined}}$, the 3×3 second rank Cartesian tensor, $B_k^{(2)}$, for each inter-nuclear vector is constructed, diagonalized into the VF, and Eqs. (21), (22), and (23) calculate each set of $\{S_{RDC}^2, \eta, \theta_{\text{avg}}^{MF}, \phi_{\text{avg}}^{MF}, \phi^l\}_k$.

ORIMUM procedure

Optimized RDC-based Iterative and Unified Model-free analysis (ORIMUM) is an iterative approach (see Fig. 2) and is related to but different from the SCRIM and iterative DIDC approaches as discussed in this section. The scheme is summarized as follows. First, alignment tensors, $\langle \mathbf{A} \rangle$, are calculated with Eq. (6) and are used to determine $\langle \mathbf{B} \rangle_{\text{refined}}$. A comparison of the effects of scaling the RDCs with $A_{zz,l}^{PAS}$ in the determination of $\langle \mathbf{B} \rangle_{\text{refined}}$ will be presented in the Applications section below [see Eqs. (24) and (25)]. Based on $\langle \mathbf{B} \rangle_{\text{refined}}$, the 3×3 symmetric Saupe matrix is constructed for the inter-nuclear vectors using expressions defined in Eqs. (19a)–(19e), and $B_k^{(2)}$ is put into the local

PAS. Utilizing Eqs. (21), (22), and (23), each set of $\{S_{RDC}^2, \eta, \theta^{MF}, \phi^{MF}, \phi'\}_k$ is extracted. These refined angles ($\theta_k^{MF}, \phi_k^{MF}$) are used as input for the next cycle of ORIUM. The cycle is finished when the convergence of order parameter is achieved using the relationship

$$\frac{1}{K} \sum_{k=1}^K |S_{RDC,k}^2(r) - S_{RDC,k}^2(r-1)| \leq 0.0001 \quad (26)$$

where r is a cycle of iteration.

The ORIUM approach differs from the SCRM method as follows. There is a minor difference: with SCRM, the inter-nuclear vector coordinates are defined in terms of spherical harmonics, while ORIUM utilizes Cartesian coordinates. The relationship between the spherical harmonics and the Cartesian coordinates are given by Eqs. (19a)–(19e). A key difference is that SCRM requires the five alignment tensor parameters $\{A_{zz}^{PAS}, R, \alpha, \beta, \gamma\}_l$ to construct the **F** matrix for the determination of $\langle \mathbf{Y} \rangle_{\text{refined}}$. DIDC and ORIUM calculate $\langle \mathbf{B} \rangle_{\text{refined}}$ directly from $\langle \mathbf{A} \rangle_{\text{normalized}}$. Finally, SCRM maximizes $\langle Y_{2,0}(\theta_k^{VF}, \phi_k^{VF}) \rangle$, whereas ORIUM places $B_k^{(2)}$ into a local axis system by diagonalization of the symmetric 3×3 second rank Cartesian tensor.

There are three key differences between ORIUM and the iterative DIDC approach. First, a grid search is implemented with the iterative DIDC scheme which minimizes the difference between the vector's coordinates obtained from $\langle \mathbf{B} \rangle_{\text{refined}}$ and an exhaustive list of (θ, ϕ) combinations to find dynamically averaged coordinates. As stated above, ORIUM diagonalizes $B_k^{(2)}$ into a local axis system in order to extract this information. The second key difference is that with the iterative DIDC scheme each inter-nuclear vector is constrained to be rigid ($S_{RDC}^2 = 1$). Only during the final iterative run is the $S_{RDC}^2 = 1$ constraint removed. ORIUM never constrains the dynamics of the inter-nuclear vectors during the iterative procedure. A final divergence between the two procedures is how flexible inter-nuclear vectors are removed from the calculation of the alignment tensors. In the iterative DIDC procedure, RDC data for an individual inter-nuclear vector is removed from the calculation of the alignment tensors if the error in the experimental and back-calculated RDCs is greater than a factor of 2. The calculation is restarted and RDC data for the next inter-nuclear vector is once again removed from the calculation if the deviation is greater than a factor of 2. This procedure is repeated until all inter-nuclear vectors fulfill the threshold for the error in experimental and back-calculated RDCs. At this point, the $S_{RDC}^2 = 1$ constraint is removed and a final iteration is performed. ORIUM removes the most flexible residues $S_{RDC}^2 \leq 0.95$, after Eq. (26) has been satisfied (see below) and then ORIUM is restarted until Eq. (26) is once again fulfilled.

As with the RDC-based model free analysis, the fundamental assumption is that the internal protein dynamics for each inter-nuclear vector is uncorrelated with fluctuations with the alignment tensor. Thus, a single average alignment tensor can be utilized for each medium. Molecular dynamics simulations indicate that this assumption is true for secondary structural elements, however **(B)** and **(A)** dynamics may be correlated for the most mobile regions of a protein (Louhivuori et al. 2006; Salvatella et al. 2008). To circumvent this potential inseparability of mobile inter-nuclear vectors and the alignment tensor fluctuations, the approach outlined in the SCRM procedure is followed (Lakomek et al. 2008). After convergence is achieved with Eq. (26), the residues that are the most mobile, as determined by fulfilling the relationship $S_{RDC}^2 \leq 0.95$, are removed from the **(A)** calculation and ORIUM is restarted with $\langle \mathbf{B} \rangle_{\text{refined}}$ from the previous iteration until Eq. (26) is once again satisfied.

The validity of ORIUM was assessed with synthetic RDC data containing a measurement error (0.3 Hz) for the 36 alignment media, which was generated using the RDC refined ubiquitin ensemble ERNST (PDB:2K0X) (Fenwick et al. 2011). The corresponding dynamic parameters (S_{RDC}^2 and η) were also calculated using the same ensemble. Using these synthetic RDC data, ORIUM was conducted and the resulting dynamic parameters have been compared with those calculated from the ensemble. The Pearson correlations of S_{RDC}^2 and η are 0.97 and 0.93, respectively.

It should be noted that the local PAS differs from the VF when $B_{zz,k}$ is a negative value, although the local PAS is usually the VF. In this case, the averaged vector orientation is actually orthogonal to the z axis of PAS. This issue can be alleviated by choosing a new axis system referred to the vector frame system (VFS), with eigenvalues ordered $B_{zz,k} \geq B_{xx,k} \geq B_{yy,k}$ instead of $|B_{zz,k}| \geq |B_{xx,k}| \geq |B_{yy,k}|$. It should also be noted that η from the VFS and the PAS can be significantly different in the case that $B_{zz,k}$ has a negative value. ORIUM utilizes the VFS after removal of residues with $S_{RDC}^2 \leq 0.95$ to obtain dynamically averaged angles of the bond vector distribution.

Determination of S_{RDC}^2 scaling factor: S_{overall}

An inherent complication when calculating S_{RDC}^2 from experimental RDCs is that dynamic averaging will reduce the actual magnitude of $A_{zz,l}^{PAS}$ or $D_{a,l}$. This reduced magnitude will result in some S_{RDC}^2 parameters over 1, and thus S_{RDC}^2 can only be considered as relative gauge of the actual amplitudes of motion, defined as $S_{RDC,unscaled}^2$ (Lakomek et al. 2006; Meiler et al. 2001). It should be noted that the alignment tensor parameters $\{R, \alpha, \beta, \gamma\}_l$ are unaffected by

the reduction in the magnitude of $A_{zz,l}^{PAS}$ or $D_{a,l}$. Two avenues to circumvent this complication have been developed. Either all the order parameters are scaled relative to the largest $S_{RDC,unscaled}^2$ leaving one order parameter equal to one (iterative DIDC approach) (Tolman 2002; Yao et al. 2008), or $S_{RDC,unscaled}^2$ is scaled relative to the Lipari-Szabo order parameters (S_{LS}^2) calculated for each residue (MFA/SCRM approach) (Lakomek et al. 2006, 2008). The problem with the first approach is that the resulting S_{RDC}^2 parameters will underestimate the amplitude of motion for each inter-nuclear vector. Overestimation can only occur if the largest $S_{RDC,unscaled}^2$ parameter has a large experimental error, leading to an artificially greater value for this parameter than in reality. Sub- and supra- τ_c motion happening for each vector equally will not be picked up by this approach, underestimating the motion except for the mentioned case. As for the second approach, S_{LS}^2 are required which may not be available for the vectors being analyzed. While this approach has been successfully applied, it may also underestimate motion since a general supra- τ_c motion affecting all the nuclei will not be picked up by this approach. Comparison of the $S_{overall}$ derived in Lakomek et al. 2008 and Lange et al. 2008 with the average order parameter from solid state data (Schanda et al. 2010) shows that the solid state NMR derived average order parameter is smaller than the one derived by this second approach suggesting that supra- τ_c motion affecting all nuclei is seen by solid state NMR but not the S_{LS}^2 versus S_{RDC}^2 approach.

Here, we present a new method for determining $S_{overall}$ without the requirement of additional information, such as S_{LS}^2 , which may not be available for the inter-nuclear vectors under investigation. The scaling procedure separates an inter-nuclear vector's motion into its principal axes in Cartesian space and leads to parameters that have a more straightforward physical interpretation. The inter-nuclear vector's motional variance is directly related to the resulting eigenvalues calculated from diagonalization of $B_k^{(2)}$ into a local axis system. The methodology outlined below exploits the fact that variance cannot be negative by definition. Therefore, a uniform scaling parameter, $S_{overall}$, is necessary to insure that the variance for each inter-nuclear vector about each of the three principal axes is positive. In the following, we present a brief outline for the derivation of bond vector motional variance for the determination of $S_{overall}$.

For each vector, the following relationships between the dynamically averaged Eigenvalues and the unit vector coordinates (x, y, z) within the VF, as shown in Eq. (4a), are as follows

$$B_{zz} = \frac{3\langle z^2 \rangle - 1}{2}, \quad B_{xx} = \frac{3\langle x^2 \rangle - 1}{2}, \quad B_{yy} = \frac{3\langle y^2 \rangle - 1}{2}. \quad (27)$$

The normalization condition sets $x^2 + y^2 + z^2 = 1$, which also implies $\langle x^2 \rangle + \langle y^2 \rangle + \langle z^2 \rangle = 1$. Therefore, B_{zz} can be recast as

$$B_{zz} = \frac{2 - 3\langle x^2 \rangle - 3\langle y^2 \rangle}{2}. \quad (28)$$

Utilizing the definitions of S_{RDC}^2 and η [Eqs. (21), and (22)], we can now reformulate S_{RDC}^2 and η in terms of the Cartesian coordinates defined within the VF

$$S_{RDC}^2 = 1 - 3\langle x^2 \rangle + 3\langle x^2 \rangle^2 - 3\langle y^2 \rangle + 3\langle x^2 \rangle \langle y^2 \rangle + 3\langle y^2 \rangle^2 \quad (29)$$

$$\eta = \frac{\sqrt{3}(\langle x^2 \rangle - \langle y^2 \rangle)}{2\sqrt{S_{RDC}^2}}. \quad (30)$$

The definition of variance is $\sigma_k^2 = \langle (k - \bar{k})^2 \rangle$, where $k = x, y$. Therefore, σ_k^2 can be substituted for $\langle k^2 \rangle$. Now, S_{RDC}^2 and η are defined in terms of variance

$$S_{RDC}^2 = 1 - 3\sigma_x^2 + 3(\sigma_x^2)^2 - 3\sigma_y^2 + 3\sigma_x^2\sigma_y^2 + 3(\sigma_y^2)^2 \quad (31)$$

$$\eta = \frac{\sqrt{3}(\sigma_x^2 - \sigma_y^2)}{2\sqrt{S_{RDC}^2}}. \quad (32)$$

Solving the system of equations gives the inverse relationships

$$\sigma_x^2 = \frac{1 - \sqrt{(1 - \eta^2)S_{RDC}^2} + \eta\sqrt{3S_{RDC}^2}}{3} \quad (33)$$

$$\sigma_y^2 = \frac{1 - \sqrt{(1 - \eta^2)S_{RDC}^2} - \eta\sqrt{3S_{RDC}^2}}{3} \quad (34)$$

A graphical depiction of the mapping between these parameters is shown in Figure S1. Using the relationship ($S_{RDC}^2 = S_{overall}^2 S_{RDC,unscaled}^2$), these equations can be written as

$$\sigma_x^2 = \frac{1 - S_{overall}\sqrt{(1 - \eta^2)S_{RDC,unscaled}^2} + S_{overall}\eta\sqrt{3S_{RDC,unscaled}^2}}{3} \quad (35)$$

$$\sigma_y^2 = \frac{1 - S_{overall}\sqrt{(1 - \eta^2)S_{RDC,unscaled}^2} - S_{overall}\eta\sqrt{3S_{RDC,unscaled}^2}}{3} \quad (36)$$

Since the variance must always be positive, the axis with the least variance (σ_y^2) should also be positive. Thus, the following inequalities are derived relating S_{RDC}^2 and η to σ_y^2

$$\sigma_y^2 = \frac{1 - S_{overall} \sqrt{(1 - \eta^2) S_{RDC,unscaled}^2} - S_{overall} \eta \sqrt{3 S_{RDC,unscaled}^2}}{3} \geq 0 \tag{37}$$

$$S_{overall} \leq \frac{1}{\sqrt{(1 - \eta^2) S_{RDC,unscaled}^2} + \eta \sqrt{3 S_{RDC,unscaled}^2}} = -\frac{1}{2B_{yy}^{unscaled}} = S_{overall}^{max} \tag{38}$$

Using Eq. (38), residue-specific $S_{overall}^{max}$ can be obtained using S_{RDC}^2 and η , or from the lowest eigenvalue. The eigenvalue definition of $S_{overall}^{max}$ follows directly from Eq. (27).

Since the reduction of magnitude in the alignment due to dynamic averaging is a global effect throughout all residues, the least residue-specific $S_{overall}^{max}$ may be utilized as the scaling factor, if there is no experimental error. The previous method in which all order parameters are scaled relative to the largest $S_{RDC,unscaled}^2$, leaving one order parameter equal to one (iterative DIDC approach) (Tolman 2002; Yao et al. 2008) is related to this new approach. If bond vector anisotropy is assumed to be axially symmetric ($\eta = 0$), $S_{overall}^{max}$ turns into $1/S_{RDC,unscaled}$. This is identical to scaling all inter-nuclear vectors such that the largest is 1 (Figure S1).

This scaling approach using the lowest residue-specific $S_{overall}^{max}$ may introduce a systematic bias due to the fact that experimental data contain errors. In order to alleviate the systematic bias, we used a statistical procedure accounting for the effect of experimental noise on $S_{overall}$ without any knowledge of S_{LS}^2 unlike the SCRM approach. First, scaling factors were calculated from the original data as well as datasets with noise added equivalent to the experimental error. These scaling factors were used to determine a value (which we term $S_{overall}^{95\%}$), below which there was a 95 % chance that the true $S_{overall}$ would fall. Given the maximum scaling factor from the original data that fulfills the constraint equation for all inter-nuclear vectors, $S_{overall}^{max}$, and corresponding set of values from noise added data (NAD), $S_{overall,NAD}^{max}$, the $S_{overall}^{95\%}$ value can be calculated as follows:

$$S_{overall}^{95\%} = \frac{S_{overall}^{max}}{\langle S_{overall,NAD}^{max} \rangle} \text{quantile}(S_{overall,NAD}^{max}, 95\%) \tag{39}$$

where the quantile function returns the given quantile of the set. The quantile prefactor compensates for systematic

shifts resulting from the addition of experimental error. With the previous study (Lakomek et al. 2008), the determination of $S_{overall}$ was conservative in order to circumvent the chance for over-estimating the supra- τ_c motion, reflected in the reported S_{RDC}^2 . Here, the criterion for scaling is that σ_y^2 should be positive, which possesses no time-scale bias. Yet, it should be noted that this overall order parameter is an upper limit for $S_{overall}$ since it could underestimate motion if there is a uniform sub- or supra- τ_c motion affecting all vectors. This is summarized in Table 1.

Applications

Ubiquitin: comparison of ORIUM with SCRM

In order to compare ORIUM with the SCRM method, N-H^N RDCs were used from measurements performed in 36 different alignment media (D36M) for the 76-residue protein ubiquitin (see Lakomek et al. 2008 for RDCs and references therein). The X-ray structure 1UBI of ubiquitin was used as the input structure for the first cycle of ORIUM (Ramage et al. 1994). For the error estimation of each extracted set of $\{S_{RDC}^2, \eta\}_k$, 1,000 Monte Carlo simulations were performed by adding uncertainty to the RDCs drawn from a Gaussian distribution with a standard deviation given by the error in the RDC set (0.3 Hz). On a single core of an Intel Core i7-2635QM CPU, the 1,000 Monte Carlo simulations required 18 min for ORIUM versus 83 min for SCRM, where the convergence criterion was set as in the SCRM implementation (Lakomek et al. 2008). When we used a 100-fold stricter convergence criterion than SCRM (see Eq. 26), the calculation was still faster on the same CPU (74 min). Thus, ORIUM is an optimized approach for the better convergence of the dynamic parameters.

A comparison of the D36M RDC data set analyzed with ORIUM and the SCRM method (re-implemented in this study) is presented in Fig. 3, which shows S_{RDC}^2 and η determined for each residue (see Table S1 for the actual values calculated from the ORIUM analysis of the D36M data set). The correlation coefficients for the S_{RDC}^2 and η parameters are both 0.99, which shows that in principle both the iterative DIDC and SCRM should yield identical results. However, in the previous implementations of DIDC (Tolman 2002; Yao et al. 2008), the effect of the alignment magnitude on the angle calculation, as shown in Eqs. (24) and (25), was not recognized, leading to variations in the S_{RDC}^2 and η values (Figure S2).

To determine whether normalization by alignment strength produces more accurate results, we compared Q factors for each alignment condition (Figure S3). For both

Table 1 Summary of the methods for determining $S_{overall}$

Scaling	$S_{RDC}^2 \leq 1$	$S_{RDC}^2 \leq S_{LS}^2$	$\sigma_y^2 \geq 0$ (ORIUM)	Solid state	3D GAF cross validation
Advantage	No other data required	Sub- τ_c motion included	No other data required	All motion is reflected in the order parameters	No other data are required
Disadvantage	Motion in the most rigid internuclear vector leads to underestimation of $S_{overall}$	Homogeneous supra- τ_c motion leads to underestimation of $S_{overall}$	Homogenous sub- and supra- τ_c motion leads to underestimation of $S_{overall}$	Different sample than solution	Gaussian fluctuations assumed, several sets of RDCs required
Reference	Yao et al. (2008)	Lakomek et al. (2008)		Schanda et al. (2010)	Salmon et al. (2009)

the standard fitting procedure and a cross validation procedure, normalization produced significantly better Q factors (with respective p values of 0.022 and 0.00033). In the unnormalized case without cross validation, stronger alignment conditions showed lower Q factors than weak conditions, indicating that they were contributing disproportionately to the fit. This lack of proportionality is also evident in an examination of the alignment tensors themselves. To estimate the degree to which the five dimensional alignment tensor space is uniformly sampled, condition numbers can be used, where a lower value indicates more uniform sampling (Peti et al. 2002). We checked the condition numbers of $\langle \mathbf{A} \rangle$ and $\langle \mathbf{A} \rangle_{normalized}$ for ORIUM implemented with Eqs. (24) and (25), respectively. Figure S4 plots the condition number versus ORIUM iteration until Eq. (26) is satisfied. For the unnormalized alignment tensors ($\langle \mathbf{A} \rangle$ using Eq. 24), the condition number finishes at 8.36, while for the normalized alignment tensors ($\langle \mathbf{A} \rangle_{normalized}$ using Eq. 25) the condition number is significantly lower finishing at 6.19. This shows that normalization of the RDCs based on $A_{zz,l}^{PAS}$ is important to adjust the contributions of strong versus weak alignments in the calculation of inter-nuclear vector orientations and dynamics. It should also be mentioned that the different values of the condition numbers do not directly indicate the reliability of the dynamics but rather the degree to which alignment space is uniformly sampled.

The slight deviation in S_{RDC}^2 between ORIUM and SCRM (Fig. 3b) originates from the $S_{overall}$, which is 0.87 from the ORIUM approach compared with the reported value of 0.89 from the SCRM report using S_{LS}^2 to calculate $S_{overall}$. Remarkably, the present approach for $S_{overall}$ determination yields scaled S_{RDC}^2 parameters that are below or within error of the S_{LS}^2 values without utilizing S_{LS}^2 in the calculation of $S_{overall}$. Due to the slightly lower $S_{overall}$ with ORIUM, the average S_{RDC}^2 for ORIUM and SCRM is 0.69 and 0.72, respectively.

In addition to starting with the IUBI structure, the ORIUM calculation was also tested with random coil input

structural coordinates and results are identical to those starting with IUBI. This worked for ORIUM and not other procedures like SCRM because at the beginning of iteration, a sizable fraction of residues used for alignment tensor calculation had their largest eigenvalue with a negative sign. For these residues, the angle used for tensor calculation was then orthogonal to the mean angle. By the time the iteration was completed, no residues had negative maximum eigenvalues. This result shows the potential utility of ORIUM in the calculation of structural parameters from random structural input and perhaps used in the refinement of conformational ensembles. Both ideas are currently under investigation.

It is also interesting to compare the results from ORIUM, specifically in regard to the calculated $S_{overall}$ of 0.87, to a recent study examining the dynamics of ubiquitin in the microcrystalline state (Schanda et al. 2010). Here, the scaling of the solid-state order parameters (S_{SS}^2) is unnecessary since the protein is not tumbling and therefore the calculated order parameters should reflect the absolute magnitude of the amplitudes of motion for each internuclear vector. The time-scale of motion embodied in S_{SS}^2 spans up to about one digit microsecond (Chevelkov et al. 2010), whereas S_{RDC}^2 encompasses motion up to about one millisecond. In principle, S_{SS}^2 is expected to be higher than S_{RDC}^2 due to the time-scale of motion embodied by the two order parameters, assuming that both the conditions for ubiquitin in microcrystalline and in solution are identical. The previously reported S_{RDC}^2 values using $S_{overall}$ of 0.89 (Lakomek et al. 2008) are on average higher than the order parameters reported for the microcrystalline state (Schanda et al. 2010). The average order parameter values of S_{RDC}^2 and S_{SS}^2 for residues 2–70 are 0.75 and 0.72, respectively. This may be due to the fact that the determination of $S_{overall}$ was done to circumvent the chance for over-estimating the supra- τ_c motion and thus the $S_{overall}$ was too conservative as described earlier. Figure S5 shows that with the approach presented here for the calculation of $S_{overall}$ without any

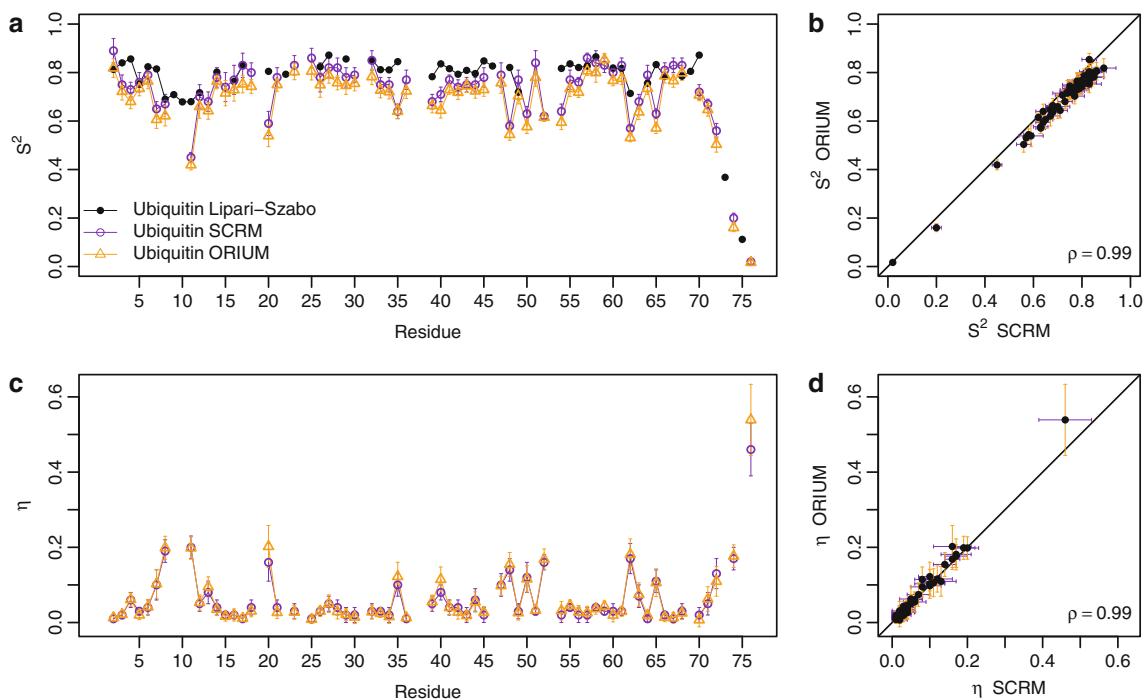


Fig. 3 Comparison of ORIUM (yellow) and SCRm (violet) derived N–H^N S^2_{RDC} and η parameters for ubiquitin. For both ORIUM and SCRm, the estimated error comes from 1,000 Monte Carlo simulations that add uncertainty to the RDCs drawn from a Gaussian distribution with a standard deviation given by the error in the RDC

set (0.3 Hz). **a** S^2_{RDC} plot by residue. The black line represents the S^2_{LS} parameters (Chang and Tjandra 2005). **b** S^2_{RDC} correlation plot. **c** η plot by residue. **d** η correlation plot

time-scale bias most of the residues possess S^2_{RDC} that are comparable or within the error of S^2_{SS} .

GB3: comparison of ORIUM with iterative DIDC

We compared ORIUM with the iterative DIDC method. N–H^N and C α –H α RDCs were used from measurements performed on 6 mutants of the third immunoglobulin domain of protein G (GB3) aligned with Pf1 phages (Yao et al. 2008). The reported error in the N–H^N and C α –H α RDCs were 0.2 and 0.4 Hz, respectively. The NMR structure 2OED of GB3 was used as the input structure for the first cycle of ORIUM (Ulmer et al. 2003). From the SECONDA analysis of the RDC data as described in the iterative DIDC publication (see Fig. 4 in Yao et al. 2008), the following RDCs were removed from the entire analysis due to inconsistencies in the structure and dynamics for these inter-nuclear vectors over the 6 different alignment media: residues 19 and 41 for the N–H^N RDCs and residues 11, 25, 30, and 40 for the C α –H α RDCs. Error estimation followed the same protocol used for ubiquitin.

The ORIUM calculation was performed with only the N–H^N RDCs or the C α –H α RDCs using the actual errors in

the RDC data of 0.2 and 0.4 Hz for the N–H^N RDCs and the C α –H α RDCs, respectively. The results are illustrated in Figs. 4 and 5 and compiled in Tables S2 and S3. The correlation for the N–H^N RDCs is 0.91 for S^2_{RDC} and is 0.92 for η . As for the C α –H α RDCs, the correlation coefficients for S^2_{RDC} and η are 0.85 and 0.77, respectively. Here, the $S_{overall}$ calculated from the ORIUM approach is 0.83 for both RDC types. This $S_{overall}$ scaling leads to an average N–H^N S^2_{RDC} of 0.65 and C α –H α S^2_{RDC} of 0.66. Thus, the S^2_{RDC} values are on average 22 % lower than in the iterative DIDC publication. As shown in Figs. 4e and 5e, iterative DIDC has significantly more negative variances than ORIUM. This is primarily because the $\sigma_y^2 \geq 0$ constraint used for ORIUM is more restrictive than the $S^2_{RDC} \leq 1$ constraint used for iterative DIDC (Figure S1), making the ORIUM S^2_{RDC} values lower than iterative DIDC.

In principle, both ORIUM and the iterative DIDC should give identical results as described earlier. The discrepancy reflected in the correlations may originate from the removal of the bias in the calculated structural and dynamic parameters due to the magnitude of alignment media [see Eqs. (24) and (25)]. In order to check this possibility, we calculated $\langle \mathbf{B} \rangle_{refined}$ with ORIUM from the N–H^N RDCs and the C α –H α RDCs simultaneously using Eq. (24) instead of Eq. (25). For

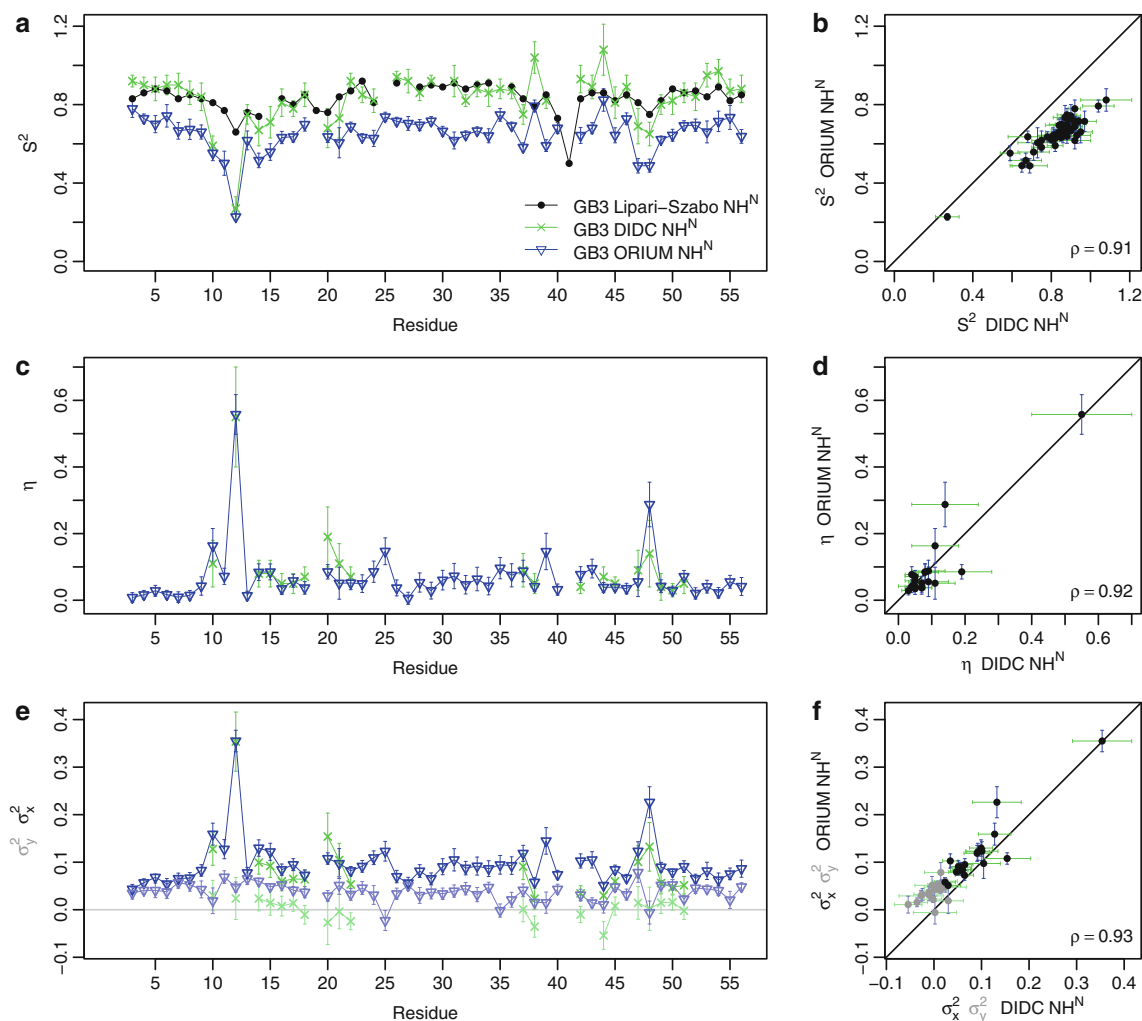


Fig. 4 Comparison of ORIUM (blue) and iterative DIDC (green) derived $N-H^N$ S^2_{RDC} and η parameters for GB3. The ORIUM calculation was performed with only the $N-H^N$ RDCs. For ORIUM, the estimated error results from 1,000 Monte Carlo simulations that add uncertainty to the RDCs drawn from a Gaussian distribution with a standard deviation given by the error in the experimental RDCs reported in the iterative DIDC publication of 0.2 Hz. From the iterative DIDC analysis, 100 Monte Carlo simulations were

performed with an uncertainty of 0.3 Hz and η is determined only for residues that were not fit to an isotropic motional model (Yao et al. 2008). **a** S^2_{RDC} plot by residue. The solid line represents the S^2_{LS} parameters (Hall and Fushman 2003). **b** S^2_{RDC} correlation plot. **c** η plot by residue. **d** η correlation plot. **e** σ_x^2 and σ_y^2 (lighter colors) by residue. **f** σ_x^2 and σ_y^2 (lighter colors) correlation plot

these calculations, uncertainties of 0.3 and 0.6 Hz were used for the MC analysis as done in iterative DIDC method and the $C\alpha-H\alpha$ RDCs were scaled by a factor of 2.08^{-1} (Bax et al. 2001) (see Figures S6 and S7). Because of the small number of alignment conditions for GB3, it is not possible to determine significant differences in a Q factor analysis with and without normalization. However, we checked the condition numbers of $\langle A \rangle$ (Peti et al. 2002) for ORIUM implemented with either Eq. (24) or Eq. (25) (see Figure S8). For the unnormalized alignment tensors, the condition number finishes at 7.84, while for the normalized alignment tensors the condition numbers are again lower finishing at an average of 7.15. The correlation for the $N-H^N$ RDCs is 0.94 for S^2_{RDC} and is 0.96 for η . As for the $C\alpha-H\alpha$ RDCs, the correlation coefficients

for S^2_{RDC} and η are 0.88 and 0.91, respectively. Although the correlations are improved, the remaining discrepancies may result from the differences in the implementation of ORIUM versus the iterative DIDC. The iterative DIDC method utilizes a grid search to find $(\theta_{avg,k}^{MF}, \phi_{avg,k}^{MF})$ for each inter-nuclear vector. During the grid search, each vector is constrained to be rigid, ($S^2_{RDC} = 1$), until the final iterative run when the constraints are lifted and the dynamic parameters calculated. As in case with normalized ORIUM, unnormalized ORIUM shows lower S^2_{RDC} values than DIDC, which results from the less restrictive DIDC constraint ($S^2_{RDC} \leq 1$) allowing more negative variances (Figures S6 and S7).

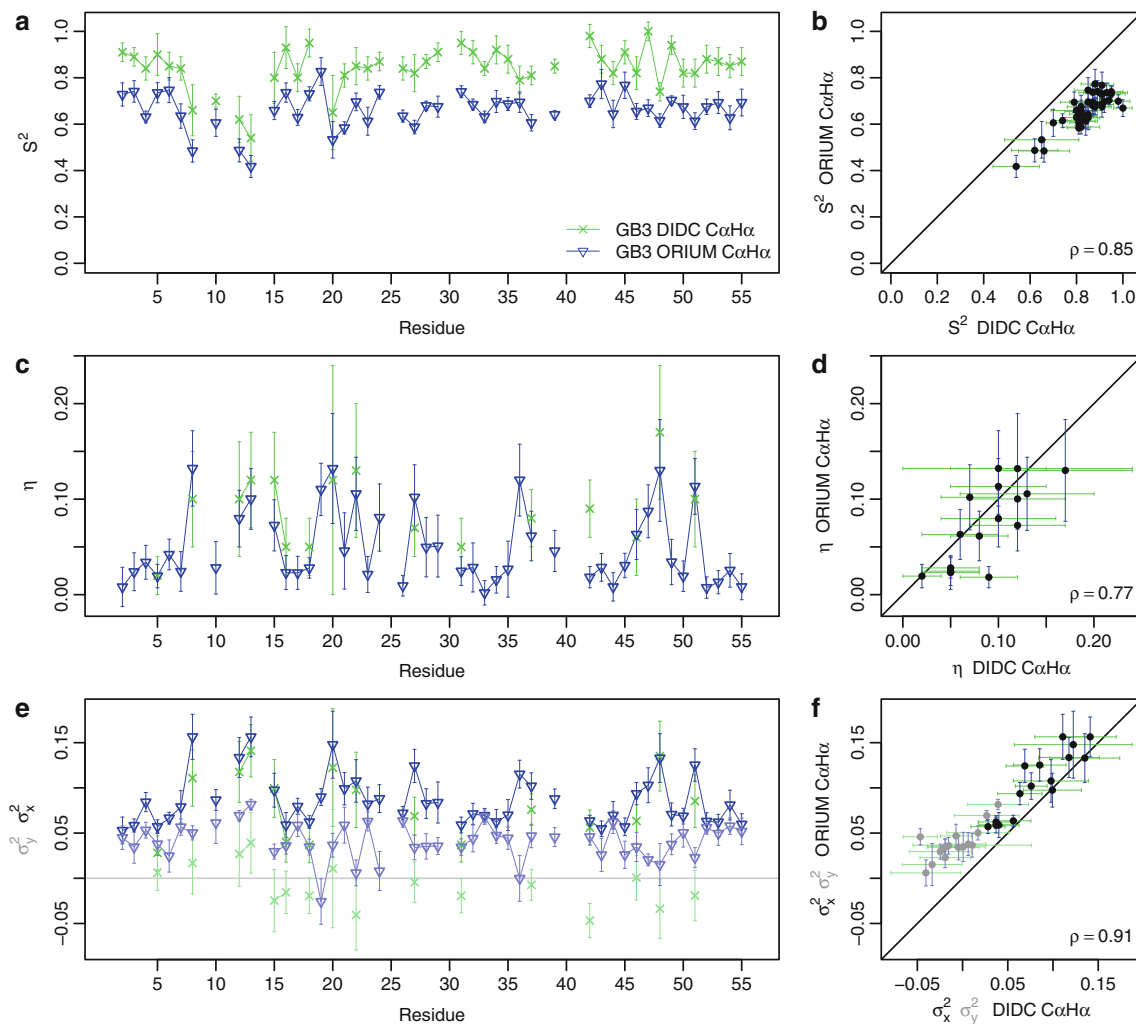


Fig. 5 Comparison of ORIUM (blue) and iterative DIDC (green) derived $C\alpha$ - $H\alpha$ S_{RDC}^2 and η parameters for GB3. The ORIUM calculation was performed with only the $C\alpha$ - $H\alpha$ RDCs. For ORIUM, the estimated error results from 1,000 Monte Carlo simulations that add uncertainty to the RDCs drawn from a Gaussian distribution with a standard deviation given by the error in the experimental RDCs reported in the iterative DIDC publication of 0.4 Hz. From the iterative DIDC analysis, 100 Monte Carlo simulations were

performed with an uncertainty of 0.6 Hz and η is determined only for residues that were not fit to an isotropic motional model (Yao et al. 2008). **a** S_{RDC}^2 plot by residue. The solid line represents the S_{LS}^2 parameters. **b** S_{RDC}^2 correlation plot. **c** η plot by residue. **d** η correlation plot. **e** σ_x^2 and σ_y^2 (lighter colors) by residue. **f** σ_x^2 and σ_y^2 (lighter colors) correlation plot

Conclusions

With ORIUM, we present a computationally efficient method for extracting structural and dynamic information for inter-nuclear vectors from RDCs unifying previously published concepts into one compact protocol. Furthermore, we demonstrate a new scheme for scaling the derived S_{RDC}^2 parameters based on variances of a single type of RDC without needing S_{LS}^2 as a constraint which constitutes an upper limit of $S_{overall}$. Dynamics occurring on time-scales slower than the rotational correlation time of proteins, encoded in RDCs, have important implications protein functionality, including

enzyme catalysis, molecular recognition, and correlated motions. The concepts set forth in this paper will go far in streamlining the procedure for calculating the dynamic average orientation and associated amplitudes of motion for inter-nuclear vectors in an iterative manner as long as RDC data sets are acquired in at least five independent alignment media.

Acknowledgments This paper is dedicated to Richard R. Ernst on the occasion of his 80th birthday. We thank Beat Vögeli and Bert L. de Groot for useful discussions. This work has been supported by the Max-Planck Society and the EU (ERC grant agreement number 233227). T.M.S and C.A.S. were supported by an Alexander von Humboldt Foundation postdoctoral research fellowship.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Ban D, Funk M, Gulich R, Egger D, Sabo TM, Walter KFA, Fenwick RB, Giller K, Pichierri F, de Groot BL, Lange OF, Grubmüller H, Salvatella X, Wolf M, Loidl A, Kree R, Becker S, Lakomek NA, Lee D, Lunkenheimer P, Griesinger C (2011) Kinetics of conformational sampling in ubiquitin. *Angew Chem Int Edit* 50:11437–11440
- Ban D, Gossert AD, Giller K, Becker S, Griesinger C, Lee D (2012) Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead. *J Magn Reson* 221:1–4
- Bax A, Kontaxis G, Tjandra N (2001) Dipolar couplings in macromolecular structure determination. *Method Enzymol* 339:127–174
- 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
- Bertini I, Del Bianco C, Gelis I, Katsaros N, Luchinat C, Parigi G, Peana M, Provenzani A, Zoroddu MA (2004) Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proc Natl Acad Sci USA* 101:6841–6846
- 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
- Chang S-L, Tjandra N (2005) Temperature dependence of protein backbone motion from carbonyl ^{13}C and amide ^{15}N NMR relaxation. *J Magn Res* 174:43–53
- Chevelkov V, Xue Y, Linsler R, Skrynnikov NR, Reif B (2010) Comparison of solid-state dipolar couplings and solution relaxation data provides insight into protein backbone dynamics. *J Am Chem Soc* 132:5015–5017
- Fenwick RB, Esteban-Martín 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
- Hall JB, Fushman D (2003) Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *J Biomol NMR* 27:261–275
- Hus JC, Brüschweiler R (2002) Principal component method for assessing structural heterogeneity across multiple alignment media. *J Biomol NMR* 24:123–132
- Hus JC, Peti W, Griesinger C, Brüschweiler R (2003) Self-consistency analysis of dipolar couplings in multiple alignments of ubiquitin. *J Am Chem Soc* 125:5596–5597
- Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by n-15 inverse detected heteronuclear NMR-spectroscopy—application to staphylococcal nuclease. *Biochemistry* 28:8972–8979
- Lakomek NA, 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
- Lakomek NA, Walter KF, Farés C, Lange OF, de Groot BL, Grubmüller H, Brüschweiler R, Munk A, Becker S, Meiler J, Griesinger C (2008) Self-consistent residual dipolar coupling based model-free analysis for the robust determination of nanosecond to microsecond protein dynamics. *J Biomol NMR* 41:139–155
- Lange OF, Lakomek NA, Farés C, Schröder GF, Walter KFA, Becker S, Meiler J, Grubmüller 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
- Lipari G, Szabo A (1982a) Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* 104:4546–4559
- Lipari G, Szabo A (1982b) Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc* 104:4559–4570
- Louhivuori M, Otten R, Lindorff-Larsen K, Annala 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, Brüschweiler R (2001) Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. *J Am Chem Soc* 123:6098–6107
- Meirovitch E, Lee D, Walter KF, Griesinger C (2012) Standard tensorial analysis of local ordering in proteins from residual dipolar couplings. *J Phys Chem B* 116:6106–6117
- Palmer AG 3rd (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev* 104:3623–3640
- Peti W, Meiler J, Brüschweiler R, Griesinger C (2002) Model-free analysis of protein backbone motion from residual dipolar couplings. *J Am Chem Soc* 124:5822–5833
- Ramage R, Green J, Muir TW, Ogunjobi OM, Love S, Shaw K (1994) Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem J* 299:151–158
- Rodriguez-Castañeda F, Haberz P, Leonov A, Griesinger C (2006) Paramagnetic tagging of diamagnetic proteins for solution NMR. *Magn Reson Chem* 44:S10–S16
- Salmon L, Bouvignies G, Markwick P, Lakomek NA, Showalter S, Li DW, Walter KFA, Griesinger C, Brüschweiler R, Blackledge M (2009) Protein conformational flexibility from structure-free analysis of NMR dipolar couplings: quantitative and absolute determination of backbone motion in ubiquitin. *Angew Chem Int Edit* 48:4154–4157
- 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
- Saupe A (1964) Kernresonanzen in Kristallinen Flüssigkeiten in Kristallinflüssigen Lösungen. I. *Z Naturforsch A* 19:161–171
- Saupe A (1968) Recent results in field of liquid crystals. *Angew Chem Int Edit* 7:97–118
- Schanda P, Meier BH, Ernst M (2010) Quantitative analysis of protein backbone dynamics in microcrystalline ubiquitin by solid-state NMR spectroscopy. *J Am Chem Soc* 132:15957–15967
- Snyder LC (1965) Analysis of nuclear magnetic resonance spectra of molecules in liquid–crystal solvents. *J Chem Phys* 43:4041–4050
- 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 (2002) A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. *J Am Chem Soc* 124:12020–12030
- Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1997) NMR evidence for slow collective motions in cyanometmyoglobin. *Nat Struct Biol* 4:292–297

Ulmer TS, Ramirez BE, Delaglio F, Bax A (2003) Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *J Am Chem Soc* 125:9179–9191

Yao L, Vögeli B, Torchia DA, Bax A (2008) Simultaneous NMR study of protein structure and dynamics using conservative mutagenesis. *J Phys Chem B* 112:6045–6056