CrossMark

ARTICLE

Information content of long-range NMR data for the characterization of conformational heterogeneity

Witold Andrałojć[1] · Konstantin Berlin[2] · David Fushman[2] · Claudio Luchinat[1,3] · Giacomo Parigi[1,3] · Enrico Ravera[1,3] · Luca Sgheri[4]

Abstract Long-range NMR data, namely residual dipolar couplings (RDCs) from external alignment and paramagnetic data, are becoming increasingly popular for the characterization of conformational heterogeneity of multidomain biomacromolecules and protein complexes. The question addressed here is how much information is contained in these averaged data. We have analyzed and compared the information content of conformationally averaged RDCs caused by steric alignment and of both RDCs and pseudocontact shifts caused by paramagnetic alignment, and found that, despite the substantial differences, they contain a similar amount of information. Furthermore, using several synthetic tests we find that both sets of data are equally good towards recovering the major state(s) in conformational distributions.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-015-9951-6) contains supplementary material, which is available to authorized users.

✉ Claudio Luchinat
luchinat@cerm.unifi.it

David Fushman
fushman@umd.edu

[1] Center for Magnetic Resonance (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

[2] Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742, USA

[3] Department of Chemistry "Ugo Schiff", University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

[4] Istituto per le Applicazioni del Calcolo, Sezione di Firenze, CNR, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy

Introduction

Biological macromolecules are inherently flexible objects and often accomplish their task through extensive conformational rearrangement (Sicheri and Kuriyan 1997; Pickford and Campbell 2004; Zhang and Zuiderweg 2004; Tonks 2006; Chuang et al. 2010). Characterization of such rearrangements and the relevant conformational states can provide important clues about the mechanisms underlying biological function. This however is a challenging task because the system is underdetermined, implying a large degeneracy in the reconstructed solutions, and requires extensive experimental work often involving multiple techniques (Bonvin and Brunger 1996; Choy and Forman-Kay 2001; Svergun et al. 2001; Burgi et al. 2001; Clore and Schwieters 2004; Schroeder et al. 2004; Iwahara et al. 2004; Bertini et al. 2004a; Blackledge 2005; Lindorff-Larsen et al. 2005; Fragai et al. 2006; Tolman and Ruan 2006; Boehr et al. 2006; Ryabov and Fushman 2006; Chen et al. 2007; Bernadò et al. 2007; Bertini et al. 2007; Ryabov and Fushman 2007; Lange et al. 2008; Hulsker et al. 2008; Korzhnev and Kay 2008; Nodet et al. 2009; Boehr et al. 2009; Stelzer et al. 2009; Huang and Grzesiek 2010; Fisher et al. 2010; Bashir et al. 2010; Rinnenthal et al. 2011; Bothe et al. 2011; Fisher and Stultz 2011; Berlin et al. 2013; Russo et al. 2013; Guerry et al. 2013; Kukic et al. 2014; Ravera et al. 2014; Torchia 2015). Therefore, it is important to know the information content provided by various experimental methods in order to decide on an optimal set of experiments a priori.

Residual dipolar couplings (RDC; Lohman and Maclean 1978) are widely used as a source of information on biomolecular structure and dynamics (Tolman 2001; Tolman and Ruan 2006; Berlin et al. 2013; Ravera et al. 2014). They arise in the presence of partial molecular orientation, which can be achieved by interactions with alignment media surrounding the molecule (Tolman et al. 1995; Tjandra and Bax 1997; Hansen et al. 1998; Losonczi and Prestegard 1998; Ramirez and Bax 1998; Wang et al. 1998; Al-Hashimi et al. 2000; Prestegard et al. 2000; Zweckstetter and Bax 2001; Lakomek et al. 2008) and/or by the preferential orientation of the molecule itself in a magnetic field due to its magnetic susceptibility anisotropy (Lohman and Maclean 1978; Tolman et al. 1995; Zhang et al. 2003; Latham et al. 2008; Ravera et al. 2014; Musiani et al. 2014). RDCs obtained by alignment induced by an external orienting medium, herein referred to as diamagnetic RDCs (dRDC), depend on the nature of the interactions of the biomolecule with the medium. These interactions can be steric and/or electrostatic and, because of this, dRDC are reporters also on the overall shape of the macromolecule and/or its charge distribution (Zweckstetter and Bax 2000; Zweckstetter 2008; Berlin et al. 2009; Maltsev et al. 2014). On the other hand, RDCs caused by molecular self-alignment, often induced by the presence of a paramagnetic center with an anisotropic magnetic susceptibility, herein termed paramagnetic RDCs (pRDC), only depend on the orientation of the internuclear vectors in the reference frame of the magnetic susceptibility tensor and are generally independent of the shape of the molecule. However, the presence of an anisotropic magnetic susceptibility also gives rise to pseudocontact shifts (PCS; Kurland and McGarvey 1970), which are reporters on the positions of the nuclei in the principal axis frame of the magnetic susceptibility tensor centered on the paramagnetic site, and therefore contain information about the structure/shape of a molecule. The use of paramagnetism-induced restraints (Gochin and Roder 1995a, b; Banci et al. 1996, 1998; Bertini et al. 2001a; Gaponenko et al. 2004; Bertini et al. 2005; Diaz-Moreno et al. 2005; Jensen et al. 2006; Bertini et al. 2008; Schmitz et al. 2012; Yagi et al. 2013b) is becoming increasingly popular because of the introduction of lanthanide binding tags (Barthelmes et al. 2011; Wöhnert et al. 2003; Rodriguez-Castañeda et al. 2006; Su et al. 2006; John and Otting 2007; Pintacuda et al. 2007; Zhuang et al. 2008; Su et al. 2008a, b; Keizers et al. 2008; Häussinger et al. 2009; Su and Otting 2010; Hass et al. 2010; Man et al. 2010; Das Gupta et al. 2011; Saio et al. 2011; Swarbrick et al. 2011a, b; Bertini et al. 2012a; Liu et al. 2012; Kobashigawa et al. 2012; Cerofolini et al. 2013; Yagi et al. 2013a; Gempf et al. 2013; Loh et al. 2013), that extend the range of applications from paramagnetic metalloproteins (Banci et al. 1996, 1997; or

proteins in which the naturally occurring metal can be replaced by a paramagnetic one; Allegrozzi et al. 2000; Bertini et al. 2001a, b, c; Bertini et al. 2003, 2004b; Balayssac et al. 2008; Bertini et al. 2010a; Luchinat et al. 2012b) to, in principle, any protein.

Given the various possibilities and limited resources, choosing the optimal set of observables for the characterization of protein conformational heterogeneity is important. In this work we analyze the information content associated with the two commonly used types of experimental data (dRDC and paramagnetic data) and discuss their features and advantages and pitfalls. Specifically, we want to understand what information can be recovered and to what extent. Importantly, the methodology that we develop below is not limited to dRDC or paramagnetic data, and can be applied to any set of experimental observables.

Theory

Formulation of the ensemble problem

We focus on analyzing the ensemble information content of three specific types of NMR restraints, dRDC, pRDC, and PCS, in the case of proteins composed of two domains connected by a flexible linker. We have used the two-domain protein calmodulin (CaM) as a test case.

As done previously (Bertini et al. 2007; Berlin et al. 2013), we assume that all three types of NMR restraints considered here represent a population-weighted average of the corresponding values for the individual conformers, and therefore have a linear dependence on the ensemble populations, such that

$$\mathbf{y} = \mathbf{a}_1 x_1 + \cdots + \mathbf{a}_N x_N + \varepsilon = \mathbf{A}\mathbf{x} + \varepsilon \quad (1)$$

where $\mathbf{y}$ is a length-L column vector representing the experimental data (dRDC, pRDC, PCS, or some combination thereof), $\mathbf{A}$ is an $L \times N$ prediction matrix consisting of N column-vectors $\mathbf{a}_j$ ($j = 1,\ldots,N$) representing the predicted data for each of the N conformers, x_j is the population weight for the jth conformer, and ε is the difference between $\mathbf{y}$ and $\mathbf{A}\mathbf{x}$ due to the presence of experimental error. This assumption seems reasonable for pRDC and PCS (Bertini et al. 2012c), whereas for dRDC the interconversion between conformers can occur on a timescale that could be comparable to the one of the interaction with the alignment medium; additionally, the latter may perturb the system.

Since in general recovering $\mathbf{x}$ from Eq. 1 is an ill-posed problem, having an infinite number of solutions, we seek to recover the minimum ensemble (sparsest solution) satisfying the experimental observables, which we express as a

constrained linear least-squares problem (Berlin et al. 2013),

$$\mathbf{x}^* = \underset{\mathbf{x}}{\arg\min} \|\mathbf{W}(\mathbf{A}\mathbf{x} - \mathbf{y})\|_2 \text{ s.t. } \mathbf{x} \geq 0, \ \|\mathbf{x}\|_0 = M \qquad (2)$$

where $\mathbf{W}$ is the weight matrix that non-uniformly weighs the residuals between $\mathbf{y}$ and $\mathbf{A}\mathbf{x}$, M is the desired ensemble size, $\|\ldots\|_2$ is the Euclidian norm, and $\|\mathbf{x}\|_0$ is the l_0 quasi-norm of $\mathbf{x}$, i.e. the number of nonzero elements in $\mathbf{x}$. Typically the experimental errors are assumed to be uncorrelated, in which case $\mathbf{W}$ is simply a diagonal matrix with $\mathbf{W}_{ii} = 1/\sigma_i$, where σ_i is the estimated experimental error of the ith observation y_i. For simplicity, for the rest of the manuscript we will drop $\mathbf{W}$ from our equations by assuming that $\mathbf{A}$ and $\mathbf{y}$ are already multiplied by $\mathbf{W}$. In the sparse ensemble selection (SES) method the ensemble size is chosen by solving the problem for reasonable values of M and using the L-curve to select the appropriate M value (Berlin et al. 2013). A different approach was also applied, based on the calculation of the maximum occurrence allowed for each conformer (MaxOcc, see below; Bertini et al. 2002a; Gardner et al. 2005; Longinetti et al. 2006; Bertini et al. 2007, 2010b, 2012b, c; Luchinat et al. 2012a; Andralojc et al. 2014).

Predicting RDC and PCS data

For steric dRDC data, we generate the prediction matrix $\mathbf{A}$ using program PATI (Berlin et al. 2009, 2013), which assumes the presence of a steric planar alignment medium (Fig. 1a). Electrostatically induced RDCs were similarly simulated using PALES (Zweckstetter 2008). The absolute scaling in the predicted dRDC values is regulated by changing the value of the parameter "liquid crystal concentration" (Zweckstetter and Bax 2000) that controls the distance between the planar steric barriers. In the SES model the absolute scaling of the predicted dRDC is treated as an implicit parameter since the sum of all weights ($\sum_j x_j$) is not constrained (Berlin et al. 2013).

For pRDC and PCS, without loss of generality, we can assume that a metal ion tag is located on the first (rigid) domain of the protein (Bertini et al. 2003). Therefore, the position of the metal ion relative to that domain is the same for all conformers. So, instead of performing the prediction of pRDC and PCS values for both domains, we obtain the prediction matrix $\mathbf{A}$ for a two-domain rigid system by first deriving the magnetic susceptibility anisotropy tensor (and metal ion's position) from the experimental data for the first domain, and then use these tensors to predict the matrix $\mathbf{A}$ values for the second domain based on its position relative to the first domain (Fig. 1b). This formulation assumes that the distribution of the relative positions of the two domains is independent of the orientation of the magnetic susceptibility anisotropy tensor in the magnetic field (Bertini et al. 2002a).

Given a specific conformer, the pRDC values in the $\mathbf{A}$ matrix are thus predicted by first deriving the vector containing the 5 independent components of the alignment tensor, $\mathbf{S}^*$, directly from the experimental data for the first domain:

$$\mathbf{S}^* = \underset{\mathbf{S}}{\arg\min} \|\mathbf{V}_1 \mathbf{S} - \mathbf{y}_1\|_2 \qquad (3)$$

where $\mathbf{V}_1$ is a 5-column matrix, the elements of which depend on the orientations of the normalized bond vectors in the fixed frame (Losonczi et al. 1999; Valafar and Prestegard 2004; Berlin et al. 2009; Simin et al. 2014) and $\mathbf{y}_1$ are the observed experimental pRDC values for the first domain. Then, using the derived $\mathbf{S}^*$, we predict the pRDC for the second domain of the jth conformer ($A_{pRDC,j}$) as

$$\mathbf{A}_{pRDC,j} = \mathbf{V}_{2j} \mathbf{S}^* \qquad (4)$$

where $\mathbf{V}_{2j}$ is the 5-column matrix of the bond vectors for the second domain in the jth conformer.

Similarly, the PCS values for the first domain can be used to derive the magnetic susceptibility anisotropy tensor $\mathbf{T}^*$, represented by a 3×3 traceless symmetric matrix, and the metal ion's position $\mathbf{p}^*$ (computed by alternating between solving a non-linear least-squares problem for $\mathbf{p}^*$, and a linear problem for $\mathbf{T}^*$). These values are then used to predict the PCS for the second domain of the jth conformer ($\mathbf{A}_{PCS,j}$). The elements of the $\mathbf{A}_{PCS,j}$ vector are the PCSs predicted for each nucleus i of the second domain, according to the relationship

$$\mathbf{A}_{PCS,j,i} \approx \frac{1}{12\pi \|\mathbf{r}_{ij}\|_2^5} \text{tr}\left(\begin{bmatrix} 3r_{ij,1}^2 - \|\mathbf{r}_{ij}\|_2^2 & 3r_{ij,1}r_{ij,2} & 3r_{ij,1}r_{ij,3} \\ 3r_{ij,1}r_{ij,2} & 3r_{ij,2}^2 - \|\mathbf{r}_{ij}\|_2^2 & 3r_{ij,2}r_{ij,3} \\ 3r_{ij,1}r_{ij,3} & 3r_{ij,2}r_{ij,3} & 3r_{ij,3}^2 - \|\mathbf{r}_{ij}\|_2^2 \end{bmatrix} \mathbf{T}^* \right) \qquad (5)$$
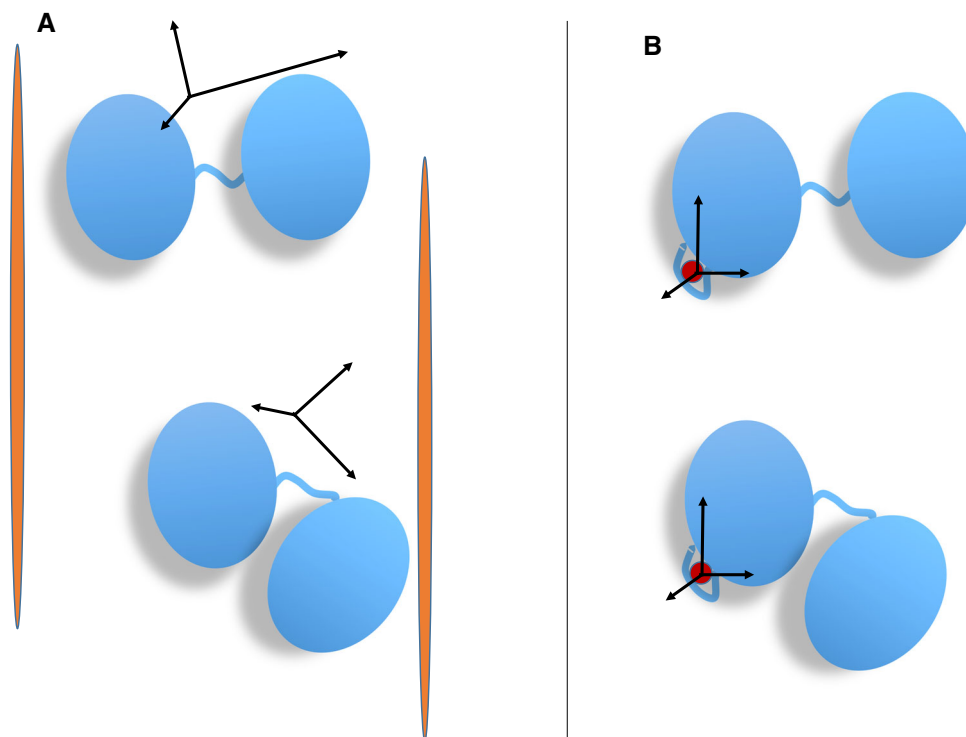
Fig. 1 Schematic illustration of the relationship between the conformation of a multidomain protein and the alignment tensor for the two experimental methods considered her (or: alignment tensor caused by external and internal alignment). In the case of partial orientation induced by external orienting media the alignment tensor changes for different conformations of a two-domain protein (**a**) whereas in the case of partial orientation induced by a paramagnetic metal ion attached to the protein the alignment tensor is invariant with respect to the orientation of the domain where it is attached (**b**)

where $\mathbf{r}_{ij} = [r_{ij,1}, r_{ij,2}, r_{ij,3}]$ is the vector connecting the metal ion (located at $\mathbf{p}^*$) and the ith atom in the jth conformer, and tr(…) designates the trace of a matrix. The elements of the tensor $\mathbf{T}^*$ and the components of the alignment tensor $\mathbf{S}^*$ are related to one another by a proportionality constant (Bertini et al. 2002b), so that each of the two can be easily calculated from the other.

Similarly to dRDC for multiple alignment media, pRDC and PCS from multiple metal ion derivatives (determined from the $\mathbf{S}^*$ and $\mathbf{T}^*$ tensors, respectively, of the corresponding metals) can be combined together in a single $\mathbf{A}$ matrix of predicted data.

Methods

Constraining SES ensemble populations

Since the scaling of the predicted dRDC values has an uncertainty (Berlin et al. 2013), when recovering SES ensembles using dRDC, we allow the total sum of $\mathbf{x}$, $\sum_j x_j$, to float, and only use the restraint $\mathbf{x} \geq 0$ (see Eq. 2).

By contrast, the values of pRDC and PCS are determined without any adjustable scaling factor, and thus the two datasets can be directly combined into a single population-constrained pRDC + PCS SES problem,

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{y}_{pRDC} \\ \mathbf{y}_{PCS} \end{bmatrix} - \begin{bmatrix} \mathbf{A}_{pRDC} \\ \mathbf{A}_{PCS} \end{bmatrix} \mathbf{x} \right\|_2 \text{ s.t. } \mathbf{x} \geq 0,$$
$$\sum_j x_j \leq c, \|\mathbf{x}\|_0 = M \tag{6}$$

where c is the upper bound on the total population weight. Since $\sum_j x_j$ represents the total population weight $\sum_j x_j$ should be 1. However, we allow for the sum of the weights to be <1, since we aim at recovering the sparsest ensemble representing the major states (potentially there could be a very large set of transient minor states). The validity of the recovered solution can be evaluated from the geometrical interpretation of pRDC: a solution is a convex combination of a set of conformers such that the averaged pRDC belong to the polyhedron with vertices in the conformers (see Figure S5; Gardner et al. 2005; Longinetti et al. 2006). Since the problem is underdetermined, there will be many solutions, and the SES method chooses to limit the number of vertices to M. In order to find a solution with this constraint, we need to use a $c < 1$ in Eq. (6). This is equivalent to shrinking the vertices of the polyhedron towards the origin by a factor c and renormalizing the weighting factors

to 1. However, since the origin is an acceptable point (Sgheri 2010a) and the set is convex, the shrunk vertices will be anyway acceptable points. In other words, if c is relatively close to 1, the conformers representing the vertices are anyway good representatives of the conformational freedom of the system. Finally, the $\sum_j x_j \leq 1$ restraint prevents from finding unphysical solutions.

SES algorithm implementation

SES ensemble recovery was implemented using the multi-orthogonal matching pursuit (MOMP) algorithm (Berlin et al. 2013). We modified the MOMP method to handle the $\sum_j x_j \leq c$ requirement using the active set method (O'Leary 2009) to restrain our solution for each iteration of MOMP. Given that there are two restraints on $\mathbf{x}$: $\mathbf{x} \geq 0$ and $\sum_j x_j \leq c$, during each iteration of the MOMP algorithm there are four possible sets of active restraints: (1) no restraints are active; (2) $\sum_j x_j \leq c$ restraint is active; (3) the $\mathbf{x} \geq 0$ restraint is active; or (4) both $\mathbf{x} \geq 0$ and $\sum_j x_j \leq c$ are active. To summarize, the constrained least-squares problem is solved as follows: update the solution using conjugate gradient (CG) method; if the solution violates $\mathbf{x} \geq 0$ or $\sum_j x_j \leq c$, solve the linearly constrained linear least-squares problem by using a "feasible direction" method (O'Leary 2009); if the solution still violates $\mathbf{x} \geq 0$, drop this solution from a list of possible solutions stored in a priority queue. This procedure is repeated for all propagated solutions from the previous iteration.

The time versus accuracy tradeoff in the MOMP algorithm is controlled by how many top solutions, K, from the current iteration are propagated to the next iteration of MOMP (Berlin et al. 2013). In order to improve the memory requirement for running SES using very large K values ($>10^6$), we modified the algorithm used to solve the overdetermined linear least-squares problem for each iteration of SES, when a new solution must be computed right after one new column is added to the list of active columns [see Supporting Information in Berlin et al. (2013)]. In the previous implementation (Berlin et al. 2013), the least-squares solution was efficiently updated by doing a rank-1 update of the QR decomposition. However, this approach requires us to store K QR decompositions during each iteration. In our current updated version, we switched to an iterative CG least-squares solver, which requires that we only store the previous-iteration solution, rather than the QR decomposition. This significantly reduced the SES memory footprint for large K. The full $\mathbf{A}^T \mathbf{A}$ matrix required for the CG algorithm is never explicitly formed, and instead the multiplication step in the CG algorithm is computed as $\mathbf{A}^T(\mathbf{A}\mathbf{x})$. With the CG implementation we are able to run SES on a 10 GB RAM desktop for $K = 10^6$, without any sacrifice in computational time or accuracy, as compared to the previous implementation.

MaxOcc calculations

The maximum occurrence (MaxOcc) of each and every conformer is defined as the maximum weight that it can obtain when part of a conformational ensemble without violating the constraints of the experimental data. No restriction is posed on the number of conformations to be included in the ensemble. Maximum occurrence (MaxOcc) can be interpreted as the maximum fraction of time that a conformation can exist, when taken together with any ensemble of conformations with optimized weights (Longinetti et al. 2006; Bertini et al. 2007; Sgheri 2010b; Bertini et al. 2010b; Das Gupta et al. 2011; Luchinat et al. 2012a; Bertini et al. 2012b, c; Cerofolini et al. 2013).

We formulate MaxOcc as a convex regularization problem, where for each conformer j we find the weight vector $\mathbf{x}$ which minimizes

$$\arg \min_{\mathbf{x}} \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda(x_j - x_{MO})^2 \right.$$
$$\left. + \lambda \left(1 - x_{MO} - \sum_{i=1, i \neq j}^{N} x_i \right)^2 \right\} \text{ s. t. } \mathbf{x} \geq 0 \quad (7)$$

where x_{MO} is the desired weight of the conformation j, and λ is a weighting factor. The calculations are repeated for increasing values of x_{MO}; the MaxOcc of conformation j is defined as the highest x_{MO} providing a value of the expression in Eq. 7 not exceeding the minimum value by more than a prefixed threshold, for example 20 %. The value of λ was fixed to 15, as found with the L-curve method, as a compromise between a good fit of the experimental observables and the proximity of the sum of the weights to 1. A frugal coordinate descent algorithm, combined with random coordinate search (Nesterov 2012), is used to solve Eq. 7.

Calculations are also performed to determine the maximum occurrence of a region (MaxOR) defined in the conformational space of the protein (Andralojc et al. 2014). The MaxOR, similar to MaxOcc, is defined as the maximum weight that a region in conformational space (composed of multiple structures) can have in an ensemble without causing a violation of the experimental restraints. First, the highest-MaxOcc structures are clustered according to their positions using a k-means algorithm as implemented in the Python library SciPy (Jones et al. 2001). The number of clusters is set to the highest value yielding reproducible clustering by the algorithm. Once the clusters are built, small regions are defined around the centers of the clusters, which include all conformations

within a given distance Δ from the center of the cluster. The MaxORs of these regions are determined by solving

$$\arg\min_{\mathrm{x}}\left\{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\left[\left(x_{MO} - \sum_{i \in C} x_i\right)^2 - \left(1 - x_{MO} - \sum_{i \in D} x_i\right)^2\right]\right\} \text{ s.t. } \mathbf{x} \geq 0 \quad (8)$$

where x_{MO} is the fixed value that must correspond to the sum of the weights of all conformations within the region, and C and D indicate the structures within and outside that region, respectively. Again, the largest x_{MO} providing a good fit of the experimental data defines the maxOR of the region.

Results and discussion

An important theoretical question that we would like to answer a priori, before performing any time-consuming simulation or experiment, is how much information for ensemble recovery is contained in dRDC versus pRDC versus PCS and in dRDC versus pRDC + PCS combined. For example, intuitively, dRDC should contain more information than pRDC, since dRDC contain shape/size-related information, while the relative informational content of PCS is harder to intuitively quantify. To what extent combining pRDC with PCS yields better results than each of these data separately? Is the information provided by pRDC + PCS similar to that provided by dRDC? Would using several different metal ions be needed to obtain results comparable to those obtained with multiple sets of dRDC, or do they produce a better set of experimental data for the characterization of the conformational heterogeneity?

In order to answer these questions, we analyzed several algebraic properties of eight experimentally feasible datasets: (1) single-alignment medium dRDC; (2) single-metal ion pRDC; (3) single-metal ion PCS; (4) single-metal ion pRDC + PCS combined; and (5–8) datasets analogous to (1–4) but with three alignment media or thee metal ions. We will refer to the one and three media/metal ions datasets as the one- and three-restraint datasets, respectively.

The datasets were generated for a pool of 32723 conformers of calmodulin (CaM), a protein composed of two rigid domains connected by a 4-residue flexible linker (Barbato et al. 1992; Tjandra et al. 1995; Chou et al. 2001; Kukic et al. 2014). This large pool of sterically allowed conformations of the protein was taken from reference (Bertini et al. 2010b), where it was generated using the program RanCh (Bernadò et al. 2007), For each conformer

and for each aligning medium or metal ion, a set of dRDCs, pRDCs, and PCSs was generated, as described in the "Theory" section.

Simulated PCS and pRDC data

The paramagnetic restraints consisted of PCS of the amide H atoms and pRDC of amide N–H pairs of the C-terminal domain of CaM induced by the presence of a paramagnetic center in its N-terminal domain. Three metals with non-coinciding magnetic susceptibility tensors (corresponding to the experimental ones obtained for Tb(III), Tm(III), and Dy(III) CaM) were used to generate three sets of PCSs (132 observations in total) and pRDCs (112 observations in total). The magnetic susceptibility anisotropy tensors were taken from reference (Bertini et al. 2009).

Simulated dRDC data

The simulated diamagnetic restraints were amide ^{15}N-^{1}H dRDCs (219 in total) induced in both CaM domains by 3 independent external alignment media: flat uncharged discs and either positively or negatively charged rods. In the first case, dRDCs were generated using PATI (Berlin et al. 2009), in the other cases using PALES (Zweckstetter and Bax 2000; Zweckstetter 2008). In both cases, the calculation of the alignment tensors, and of the corresponding dRDC, are performed under the assumption that the protein's conformations are rigid during the time course of its interaction with the alignment medium. As a word of caution we note that every interaction of a protein with the alignment medium might actually perturb its conformation, and these interactions can occur on a timescale that is slower than the conformational averaging itself. The assumption that the averaged dRDCs correspond to a weighted average of the RDCs calculated for the individual conformations, although universally used, might fall short in representing the real physical picture.

SVD of prediction matrices

The first and simplest analysis we performed was aimed at evaluating the theoretical information content of the eight different datasets described above. This was done through the spectral analysis of the prediction matrix $\mathbf{A}$ for each dataset. The spectral analysis measures the number of significant linearly independent components present in the data, by counting the eigenvalues corresponding to linearly independent eigenvectors. This directly provides an upper bound on the number of independent conformers we can hope to extract. Trying to recover a larger number of
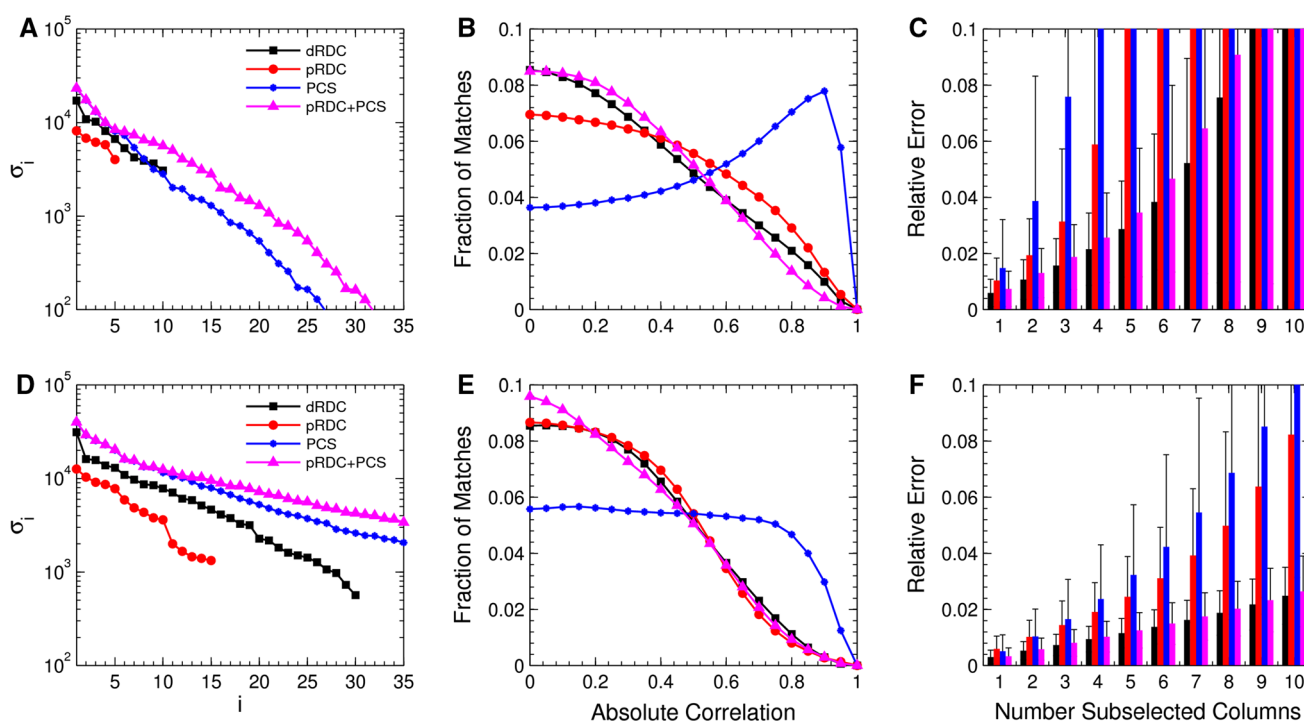
Fig. 2 SVD decomposition (*left panels*), histogram of column correlations (*center panels*), and condition number of randomly subselected set of columns (*right panels*), for the eight described datasets. The results for a single medium/metal ion are shown on the *top*, and the results for the 3 media/metal ions are shown on the *bottom*. **a–d** The 35 largest singular values of the associated **A** matrices. **b–e** The distribution of the uncentered correlations between all pairs of columns in the **A** matrix, estimated by performing 20,000 random samples. **c–f** The expected mean and SD of the relative error for recovering population weights from an arbitrary $M = 1,\ldots,10$ subset of columns

independent conformers would result in overfitting. The results are shown in Fig. 2a, d.

As shown in Eq. 3, any vector of RDC values (either pRDC or dRDC) from a rigid domain can be expressed as a matrix **V**, which can be determined from the orientations of the bond vectors of that domain, multiplied by the 5 independent components of the alignment tensor matrix. Since there is a linear dependence of the observed data on the 5 components of the alignment tensor, we expect the **A** matrix for dRDC to have rank 10 (5 independent parameters for each of the two domains), and for pRDC to have rank 5, since only the second domain data are used for ensemble recovery. The number of unknowns in the paramagnetic case is also smaller because the alignment tensor for the first domain (5 parameters) can be easily determined from PCS and pRDC measured for this domain, as they are not averaged by conformational variability.

Numerical spectral analyses of the generated prediction matrices for dRDC and pRDC (Fig. 2a, d) support our theoretical analysis, and show that the number of singular values of matrix **A** for one-restraint dRDC and pRDC data is 10 and 5, respectively. Going from 1 to 3 alignment tensors triples the number of non-zero singular values for dRDC and pRDC, as would be expected for linearly

independent alignments. The large decrease in the magnitude of singular values for the last 10 dRDC and 5 pRDC non-zero singular values in the three-restraint datasets likely reflects the difficulty in experimentally obtaining three fully independent alignment tensors. The larger magnitudes of dRDC singular values compared to the singular values for pRDC are not related to their information content, but merely reflect the relative strength of diamagnetic versus paramagnetic alignment in the simulated data. On the contrary, it is the decrease in the relative magnitude of the singular values with respect to the largest value, calculated from a set of data, that reflects the difficulty in exploiting the associated restraints, and is hence ultimately related to the information content.

Similarly, the observed PCS data for a rigid domain which is not containing the paramagnetic ion (i.e. for the second domain) can be expressed using 8 parameters: the 5 independent components/parameters defining the **T** tensor, and the 3 parameters describing the metal-ion's position **p** with respect to this domain. However, since the observed PCS vector **y** is not linearly related to **p**, the rank of $\mathbf{A}_{PCS}$ (calculated from the PCSs in the second domain) is much higher than 8, and greater than that for dRDC or pRDC datasets. The rank of $\mathbf{A}_{PCS}$ is actually close to (up to) the

number of observations; however as Fig. 2a, d show, the magnitude of the singular values decreases very rapidly. This decrease reflects the strong difference in the PCS values between conformers where the C-terminal (second) domain is close to the metal ion (paramagnetic center) and those where it is far away. After the first ≈ 15 entries, in the one-restraint case the singular values are very small because similar PCS values are calculated for conformers not very far from one another and for nuclei which are spatially close to several other nuclei. When using three sets of metal ions, the number of conformers with large and different PCS values increases. Thus, the decrease in the magnitude of the singular values is significantly slower than in the case of a single metal ion (Fig. 2a–d).

One major advantage of using metal ions instead of steric alignment is that both pRDC and PCS are collected from the same biochemical construct. Thus, two independent datasets can be directly combined, as described in Eq. 6. When combining these datasets, a significantly slower decay in singular values of **A** is obtained compared to the pRDC and PCS datasets analyzed independently. This supports the accepted intuition that pRDC and PCS provide orthogonal structural restraints (pRDCs are very sensitive to orientation, PCSs mostly provide distance restraints).

Histograms of prediction matrices

The spectral analysis of the **A** matrices suggests that pRDC + PCS and even PCS alone provide better restraints for ensemble selection than dRDC. However, singular values are not an exhaustive description of the overall vector distribution. Therefore, we directly analyzed the distribution of correlations between all columns of the matrix **A** calculated for dRDC, pRDC, and PCS. The uncentered correlation distributions between all pairs of columns are shown in Fig. 2b, e. The more uncorrelated the columns of each specific **A** ($\mathbf{A}_{dRDC}$, $\mathbf{A}_{pRDC}$, $\mathbf{A}_{PCS}$) the smaller the chance that an alternative conformer can explain the same subset of experimental data, thus decreasing the number of viable alternative ensembles. In the optimal case, all columns would have zero correlation, and the ensemble solution would be unique.

Figure 2b, e clearly demonstrate that even though the number of singular values of PCS is larger than that of dRDC and pRDC, the correlation distribution is actually significantly worse than for any other dataset, so that their information content could not be larger. The higher correlation for large fraction of the conformers reflects a distribution of PCS where very large changes occur in proximity of the metal ion only, whereas almost no change

occurs far away from the metal ion. Additional metal ions can significantly improve the distribution of correlations, although it remains poor with respect to that of the other restraints.

Since pRDCs are distance-independent, they provide a more uniform distribution of values, so that their correlation distribution is much better than for PCS. The pRDC distribution is anyway worse than that of dRDC in the one-restraint case; it significantly improves, essentially to the level of dRDC, in the three-restraint case. Interestingly, the dRDC distribution changes only slightly between one and three restraints, which suggests that the information contained in the additional dRDC datasets is more redundant than in the pRDC case.

Combining pRDC with PCS results in a better correlation distribution than for pRDC and PCS individually. In turn, the correlation distribution of pRDC + PCS is very similar to that of dRDC in the one-restraint case and actually somewhat better in the three-restraint case.

Expected relative error

While the correlation plots in Fig. 2b, e provide an estimate of the **A** matrix column vector distribution, they do not directly tell how well ensembles greater than two can be recovered, nor do they take signal-to-noise ratio into account. To assess how well larger ensembles can be recovered, we computed the mean and standard deviation (SD) of the relative error from a synthetically generated **y** data (with added Gaussian error) for $M = 1,…,10$ columns. The mean and SD were computed by randomly sampling, for each M value, M columns and uniformly at random generating the associated population weights **x**. The synthetic **y** was generated as $\mathbf{y} = \mathbf{Ax} + N(0,1)$, where $N(0,1)$ is the zero-mean Gaussian distribution with $\sigma = 1$. The vector $\mathbf{x}^*$ and the associated relative error, $\|\mathbf{x} - \mathbf{x}^*\|_2/\|\mathbf{x}\|_2$, were recovered by solving Eqs. 2 and 6. In order to guarantee a <0.1 % relative error with >99.999 % confidence using Chernoff bound, the process was repeated 40,000 times for each M. The results for all datasets are shown in Fig. 2c, f.

For the one-restraint datasets, dRDC has lower relative error than pRDC, PCS, or pRDC + PCS. As expected, there is a rapid growth in pRDC errors due to the low matrix rank, and high errors overall in PCS due to the high correlation between columns. In the case of the three-restraint datasets, dRDC has significantly lower relative error than pRDC, even though on the correlation plot the two distributions are very similar. Interestingly, combining pRDC + PCS yields only slightly higher error rate than for dRDC.

Recovering the conformational variability from synthetic datasets

In the previous sections we theoretically analyzed the information content of 8 datasets of synthetic dRDC, pRDC and/or PCS data. Here we perform a direct comparison of the performance of the different restraints in recovering information on the structural variability of the system. To achieve this, we determined (1) the minimum-size sparsest ensemble solution using the SES method (Berlin et al. 2013) and (2) the conformations (as well as the regions in the conformational space) with the highest MaxOcc values. In this way it becomes possible to analyze the accuracy of the recovered solutions from the different sets of synthetic averaged data.

For this purpose, we devised three simulations modeling (1) extensive mobility around a single conformation, (2) two-site exchange with limited mobility around each center, and iii) two-site exchange with a reduced difference in the orientations of the two centers. In each of the simulations, the two-domain protein CaM was allowed to sample different, well defined, parts of its sterically allowed conformational space. Synthetic restraints were calculated as weighted averages over the values of dRDC, pRDC, and PCS of the individual conformations belonging to the sampled regions. These average data were perturbed with a Gaussian error with a SD of 1, 2, or 3 Hz for pRDC and dRDC and of 0.01, 0.02, or 0.03 ppm for PCS.

In the following descriptions of the simulated conformational ensembles, the N-terminal domain of CaM is taken as the frame of reference, and each conformation is described by the different position and orientation of the C-terminal domain with respect to the N-terminal domain. The exact details of each simulation, although described accurately for completeness, are not crucial for the success of the ensemble recovery attempts.

Simulation 1

In this first simulation we consider the case of conformational variability centered at a single extended conformation of CaM. The sampled ensemble consists of all the conformers, present in the pool of the 32723 sterically allowed conformers, within a distance Δ (measured as a combination of translation and rotation) from the central extended structure (Fig. 3a; Bertini et al. 2012b). Specifically, this distance is defined as:

$$\Delta = d + f(1 - \cos\alpha) \qquad (9)$$

where d is the translation of the center of mass of the C-terminal domain from the central structure, and α is the angle of rotation from the central structure, calculated as $\alpha = \arccos(|q_c \cdot q|)$, where q_c and q are the unitary
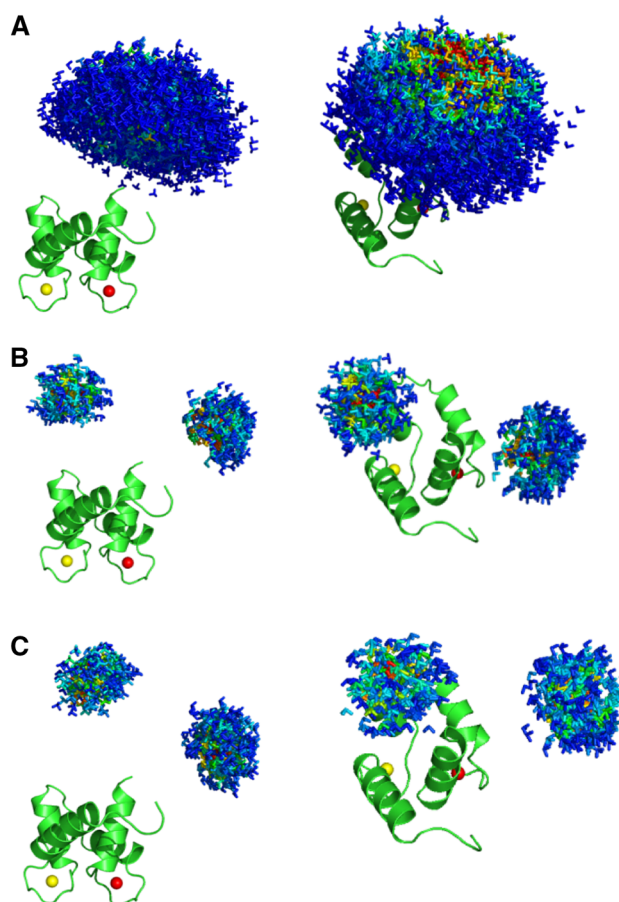

Fig. 3 The simulated ensembles. Different positions of the C-terminal domain of CaM are represented by a triad of Cartesian axes, centered at the center of mass of the C-terminal domain. The conformers are color coded according to their relative weights (form *red* = high weight to *blue* = low weight). **a** Simulation 1, **b** Simulation 2, **c** Simulation 3. The ensembles are shown from two different points of view in the *left* and *right* panels. All the conformers are superimposed by the N-terminal domain, which is shown in cartoon representation

quaternions describing the central structure and the other structure. Note that the two structures are actually 2α apart in Cartesian space (Kuffner 2004). Δ defines the largest allowed spatial displacement (when α is 0) and the largest allowed rotation (when d is 0; it also depends on the factor f) from the position of the central conformer. In the present simulation, conformations with Δ up to 30 Å (a reasonable estimate for this system) were accepted and the value of f was set to 84 Å. In this way, the conformers in the constructed ensemble can have the center of mass of the C-terminal domain at a maximum distance of 30 Å with respect to the conformer at the center of the distribution, if they have the same orientation (the distance decreases with increasing the difference in the orientation). Their C-terminal domain can be rotated up to 100° ($\alpha = 50°$) with respect to the central conformer, if there is no translation of

the center of mass (and gradually less and less as the translational component increases). The weight of each conformation in the ensemble depends on its Δ, and is fixed according to a Gaussian distribution centered at $\Delta = 0$, with SD chosen to provide weights close to zero when Δ is close to 30 Å.

Simulation 2

This simulation models the case of a two-site exchange, with limited mobility allowed around each of the two main conformers (Fig. 3b). The two centers were separated by approximately 30 Å and their C-terminal domains were rotated by ca. 140° with respect to each other. The mobility around each center was simulated as in the previous case with the threshold on Δ set to 10 Å and f equal to 42.7 Å, which corresponds to a maximum allowed angular displacement with respect to the central conformer of 80° ($\alpha = 40°$).

Simulation 3

This simulation is similar to Simulation 2, with the difference that the angular distance between the two sites was decreased almost twofold (Fig. 3c). Sites with more similar orientations are likely to present a bigger challenge in ensemble recovery using restraints which depend on the domain orientations. The distance between the centers (both distinct from those used in Simulation 2) is 30 Å while the difference in orientation of the C-terminal domains is now 80°. The threshold of Δ and the value of f used to simulate the residual mobility around each center

were the same as in Simulation 2, hence the same upper limit on the angle α.

SES ensembles

We applied the SES method to these simulated datasets and analyzed how the various restraints affect the recovery of the main conformations contained in the synthetic ensembles used to generate the data. The recovered ensembles were evaluated in terms of their sizes (number of major states) and of the proximity of recovered structures to the centers of the synthetic ensembles (in terms of spatial and angular displacement). As already mentioned in the "Theory" section, the ensemble size was chosen using the L-curve method (Berlin et al. 2013; see Figure S6).

The results are presented in Tables 1, 2 and 3. In general, dRDCs allowed a reasonably accurate recovery of the major states that were used to generate the synthetic datasets (see, for instance, Fig. 4a). However, in all three simulations, in some solutions one additional conformer was recovered, albeit with a relatively low weight. This additional conformer either belongs to the distribution of conformers around one of the main centers (as in Simulation 1 with error of 1 and 2 Hz, and in Simulation 2 with error of 2 Hz, Fig. 4b) or is positioned in-between the two major states (as in Simulation 2 with error of 1 Hz, Fig. 4c). In the first case its presence may reflect conformational heterogeneity; in the second case it is likely related to artifacts. The latter may arise because, 'average conformers' can be more compatible with the averaged experimental observables than any of the actually sampled conformations taken individually.

Table 1 Results of Simulation 1

Simulation 1		Ensemble				Conformer 1		Conformer 2	
Restraint	Error	Ensemble size	Total weight	Weight conformer 1	Weight conformer 2	Translation (Å)	Rotation (°)	Translation (Å)	Rotation (°)
PCS	0.01 ppm	1	0.792	0.792	–	3.85	15.6	–	–
	0.02 ppm	1	0.799	0.799	–	3.85	15.6	–	–
	0.03 ppm	1	0.771	0.771	–	3.85	15.6	–	–
dRDC	1 Hz	2	0.866	0.663	0.203	3.85	15.6	12.78	107.3
	2 Hz	2	0.839	0.482	0.356	8.68	22.4	6.34	39.5
	3 Hz	1	0.755	0.755	–	3.85	15.6	–	–
pRDC	1 Hz	1	0.598	0.598	–	10.48	11.8	–	–
	2 Hz	1	0.597	0.597	–	10.48	11.8	–	–
	3 Hz	1	0.608	0.608	–	14.01	15.6	–	–
pRDC + PCS	1 Hz/0.01 ppm	1	0.755	0.755	–	3.85	15.6	–	–
	2 Hz/0.02 ppm	1	0.761	0.761	–	3.85	15.6	–	–
	3 Hz/0.03 ppm	1	0.742	0.742	–	3.85	15.6	–	–

The table reports the sizes of the recovered ensembles, the specific weights ascribed to their constituent conformers, and for each of these conformers their spatial and angular displacement from the center of the original ensemble

Table 2 Results of Simulation 2

| Restraint | Error | Ensemble | | | | | Conformer 1 | | | Conformer 2 | | | Conformer 3 | | |
		Ensemble size	Total weight	Weight conformer 1	Weight conformer 2	Weight conformer 3	Closest center	Translation (Å)	Rotation (°)	Closest center	Translation (Å)	Rotation (°)	Closest center	Translation (Å)	Rotation (°)
PCS	0.01 ppm	2	1.000	0.617	0.383	–	B	3.65	26.2	A	0.00	0.0	–	–	–
	0.02 ppm	2	1.000	0.594	0.406	–	B	3.65	26.2	A	0.00	0.0	–	–	–
	0.03 ppm	2	1.000	0.557	0.443	–	A	5.44	16.8	B	3.67	42.6	–	–	–
dRDC	1 Hz	3	0.754	0.345	0.272	0.137	A	0.00	0.0	B	3.99	31.7	B	14.99	63.4
	2 Hz	3	0.881	0.423	0.328	0.130	A	3.12	8.6	B	3.67	42.6	B	4.17	37.5
	3 Hz	2	0.737	0.384	0.353	–	A	6.49	14.4	B	5.16	19.7	–	–	–
pRDC	1 Hz	2	0.674	0.338	0.336	–	A	8.45	14.1	B	14.52	21.6	–	–	–
	2 Hz	2	0.636	0.360	0.275	–	A	12.22	32.8	B	14.86	13.2	–	–	–
	3 Hz	2	0.692	0.365	0.327	–	B	9.66	22.0	A	15.56	14.6	–	–	–
pRDC + PCS	1 Hz/0.01 ppm	2	0.741	0.384	0.357	–	B	7.72	34.1	A	0.00	0.0	–	–	–
	2 Hz/0.02 ppm	2	0.926	0.481	0.445	–	B	7.85	24.3	A	0.00	0.0	–	–	–
	3 Hz/0.03 ppm	2	1.000	0.546	0.454	–	B	7.85	24.3	A	7.63	12.0	–	–	–

The table reports the sizes of the recovered ensembles, the specific weights ascribed to their constituent conformers and for each of these conformers indicates the closest of the two original sites together their spatial and angular displacement from it

Table 3 Results of Simulation 3

| Restraint | Error | Ensemble | | | | | Conformer 1 | | | Conformer 2 | | | Conformer 3 | | |
		Ensemble size	Total weight	Weight conformer 1	Weight conformer 2	Weight conformer 3	Closest center	Translation (Å)	Rotation (°)	Closest center	Translation (Å)	Rotation (°)	Closest center	Translation (Å)	Rotation (°)
PCS	0.01 ppm	1	1.000	1.000	–	–	A	10.92	24.8	–	–	–	–	–	–
	0.02 ppm	1	1.000	1.000	–	–	A	10.92	24.8	–	–	–	–	–	–
	0.03 ppm	1	1.000	1.000	–	–	A	10.92	24.8	–	–	–	–	–	–
dRDC	1 Hz	3	0.965	0.451	0.378	0.137	A	2.15	19.6	B	4.05	19.6	B	9.31	86.9
	2 Hz	2	0.820	0.431	0.389	–	A	7.09	21.6	B	4.25	24.3	–	–	–
	3 Hz	2	0.876	0.440	0.436	–	A	3.29	18.1	B	4.05	19.6	–	–	–
pRDC	1 Hz	2	0.739	0.374	0.364	–	B	9.60	20.5	A	12.36	11.0	–	–	–
	2 Hz	2	0.697	0.361	0.336	–	B	10.62	11.1	A	9.38	14.2	–	–	–
	3 Hz	2	0.720	0.362	0.358	–	B	4.25	24.3	A	5.20	13.9	–	–	–
pRDC + PCS	1 Hz/0.01 ppm	2	0.772	0.391	0.380	–	A	3.29	18.1	B	6.70	8.9	–	–	–
	2 Hz/0.02 ppm	2	0.799	0.408	0.392	–	B	17.82	19.9	A	8.03	20.0	–	–	–
	3 Hz/0.03 ppm	2	0.823	0.445	0.377	–	A	3.05	11.6	B	6.70	8.9	–	–	–

The table reports the sizes of the recovered ensembles, the specific weights ascribed to their constituent conformers and for each of these conformers indicates the closest of the two original sites together their spatial and angular displacement from it
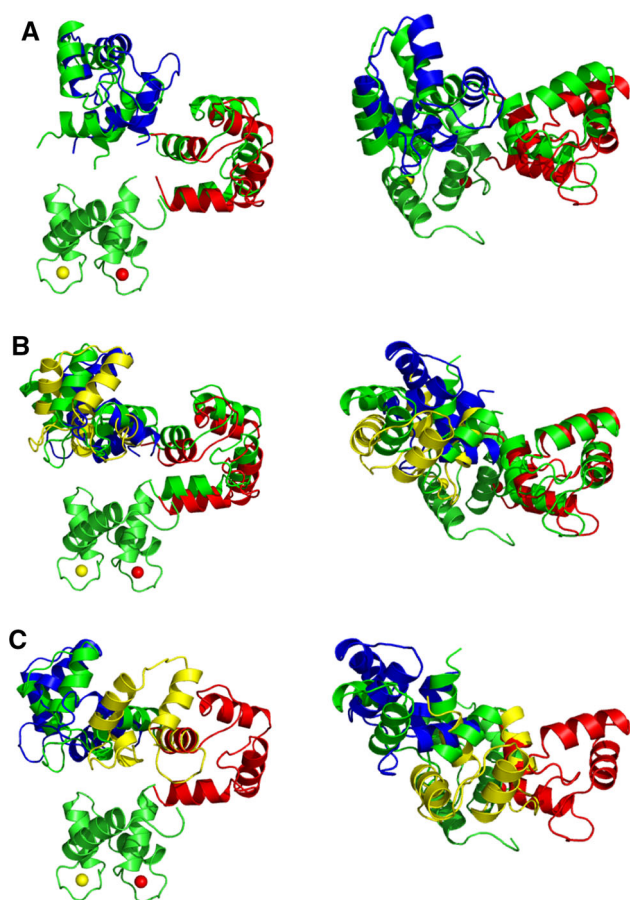
Fig. 4 SES recovery using dRDC data. Color code for the C-terminal domain: *green*—simulated conformers in the centers of the regions, *red*, *blue*, and *yellow*—reconstructed conformers with highest, intermediate, and lowest weight, respectively. **a** An ensemble with correctly recovered major states (Simulation 3, 3 Hz error), **b** An ensemble with an additional state present next to one of the centers (Simulation 2, 2 Hz error), **c** An ensemble with an additional state (*yellow*) recovered (Simulation 2, 1 Hz error). The ensembles are shown from two different points of view in the *left* and *right panels*. All conformers are superimposed by the N-terminal domain

In the case of pRDCs, the right number of major states was always recovered (Fig. S1), and in the corresponding conformers the domains were oriented with an accuracy comparable to that achieved with dRDC. It should be recalled that pRDCs contain no information whatsoever on the relative positions of the domains, which therefore results in inaccuracy of their positioning.

PCS data alone in two out of three simulations were sufficient to recover the correct solutions (Fig. S2) in terms of ensemble sizes and locations of the major states (with the accuracy similar to dRDC). However in Simulation 3, where the two states are more alike to one another, the calculations provide only a single state (Fig. S2B) situated in-between the two actual centers (in terms of both translation and orientation). The recovery of such an incorrect state is most likely, as already mentioned for dRDC, the
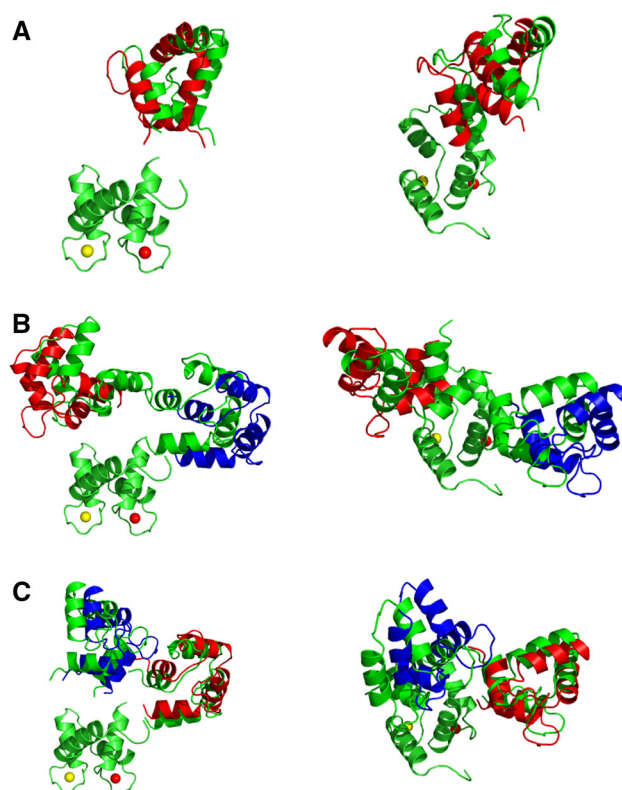

Fig. 5 SES recovery with all the paramagnetic data. Color code for the C-terminal domain: *green*—simulated conformers in the centers of the regions, *red*, *blue*—reconstructed conformers with higher and lower weight, respectively. **a** Simulation 1, **b** Simulation 2, **c** Simulation 3, with 0.03 ppm and 3 Hz errors. The ensembles are shown from two different points of view in the *left* and *right panels*. All conformers are superimposed by the N-terminal domain

outcome of the averaging of the experimental observables. Using all the paramagnetic data together (i.e. pRDC and PCS) improved the robustness of the recovery: both translations and orientations were satisfactory accurate in all cases (Fig. 5). The translation and rotation with respect to the conformers at the center of the distributions were within 4 Å and 16° for Simulation 1, 8 Å and 34° for Simulation 2, and 3 Å and 18° or 7 Å and 9° for Simulation 3 (1 Hz and 0.01 ppm error case). The ensemble recovery is robust, as increased errors did not noticeably affect the accuracy of solutions.

In conclusion, diamagnetic RDC, as well as the combination of paramagnetic RDC and PCS, are both equally suitable restraints for the recovery of the major states present in conformational ensembles. Special attention should be paid to the fact that, occasionally, 'average conformers' may be recovered.

MaxOcc analysis

Similar to the SES analysis, we performed MaxOcc analysis on the same datasets. From the MaxOcc values, it is

possible to determine which conformers can be sampled with the largest weights. In order to speed the computational analysis up, we used random sampling to detect regions of with potentially high MaxOcc conformers, and then expanded those regions, to find the globally best solution. To do this, we first computed MaxOcc for 400 conformers, randomly chosen from the generated pool (Bertini et al. 2010b, 2012b; Cerofolini et al. 2013). Then the conformers with the highest MaxOcc (up to 0.8 of the MaxOcc of the highest scoring conformer) were selected and the MaxOcc of their neighboring conformers (in the conformational space) were calculated. The procedure was repeated until no more neighbors with high MaxOcc were found. The neighboring conformers scored at each iteration were chosen using Eq. 9 with the threshold on Δ of 5 Å and $f = 40$ Å. If the final distribution of the highest MaxOcc conformers was broad, the analysis was supplemented by the maximum occurrence of regions (MaxOR) approach, which permitted to discriminate between the cases of high MaxOcc conformers corresponding to conformers actually sampled by the protein and the cases of high MaxOcc conformers corresponding to conformers arising from data averaging (Andralojc et al. 2014).

The results of the MaxOcc analysis are reported in Table 4 for all three simulations. In Simulation 1, for both the paramagnetic and diamagnetic data, the analysis revealed that all the conformers with the highest MaxOcc (from 0.8 to 1 of the highest MaxOcc, corresponding to 0.58–0.73 for the paramagnetic data and 0.57–0.71 for the dRDC) form a single, relatively compact, region in the conformational space (Fig. 6a, c). In order to quantify its

agreement with the original distribution, the center of the region was calculated by averaging the translational and orientation parameters of the highest MaxOcc conformers. The conformation so obtained was then compared with the conformation at the center of the original distribution. As shown in Table 4 and Fig. 6b, d, the agreement was very good in terms of spatial and angular displacement for both the diamagnetic and the paramagnetic data, either for 1 Hz/ 0.01 ppm or for 3 Hz/0.03 ppm errors.

In simulation 2, i.e. the case of two well separated conformational regions, when dRDC are used, the highest MaxOcc conformers are positioned in two distinct, clearly separated regions (Fig. 7a), the centers of which are positioned very close to the centers of the actually sampled distribution (Table 4; Fig. 7b). When paramagnetic data (PCS + pRDC) are used, the highest MaxOcc (0.41–0.51) conformers are positioned in one elongated, banana-shape region in the conformational space (Fig. 8a), which includes the two actually sampled centers, but also many conformers situated between them (their high score is an outcome of conformational averaging as described in the SES results paragraph). From these results, one cannot conclude whether the studied conformational ensemble mainly reflects a two-site exchange case or the sampling of all the conformations within the determined region. In order to distinguish between these two cases, MaxOR calculations were performed. The highest MaxOcc conformers were clustered in 5 regions, shown in Fig. 8b, which include all conformations with distance $\Delta \leq 5$ Å from the central conformation (calculated using eq. 9, with $f = 147$ Å). The MaxOR values for these regions are

Table 4 The MaxOcc/MaxOR analysis

Simulation	Restraint	Error	Recovered center 1			Recovered center 2		
			Closest center	Translation (Å)	Rotation (°)	Closest center	Translation (Å)	Rotation (°)
Simulation 1	dRDC	1 Hz	–	3.42	5.7	–	–	–
		3 Hz	–	2.65	10.3			
	pRDC + PCS	1 Hz/0.01 ppm	–	4.72	9.1	–	–	–
		3 Hz/0.03 ppm	–	4.16	9.5	–	–	–
Simulation 2	dRDC	1 Hz	A	3.44	4.4	B	1.71	25.5
		3 Hz	A	3.27	9.7	B	3.26	29.8
	pRDC + PCS	1 Hz/0.01 ppm*	A	4.42	14.4	B	1.02	28.1
		3 Hz/0.03 ppm	A	7.15	25.1	B	6.60	24.7
Simulation 3	dRDC	1 Hz*	A	0.99	14.3	B	8.17	16.6
		3 Hz	A	5.32	28.9	B	5.35	16.0
	pRDC + PCS	1 Hz/0.01 ppm*	A	4.39	10.8	B	3.57	31.4
		3 Hz/0.03 ppm	A	5.45	9.9	B	5.80	40.3

For each simulation the spatial and angular displacement of the center of the ensemble of the best scoring conformers from the center of the actually sampled distribution is reported, together with the indication of the closest site, if applicable. The cases where MaxOR was used are indicated with an asterisk next to the error level
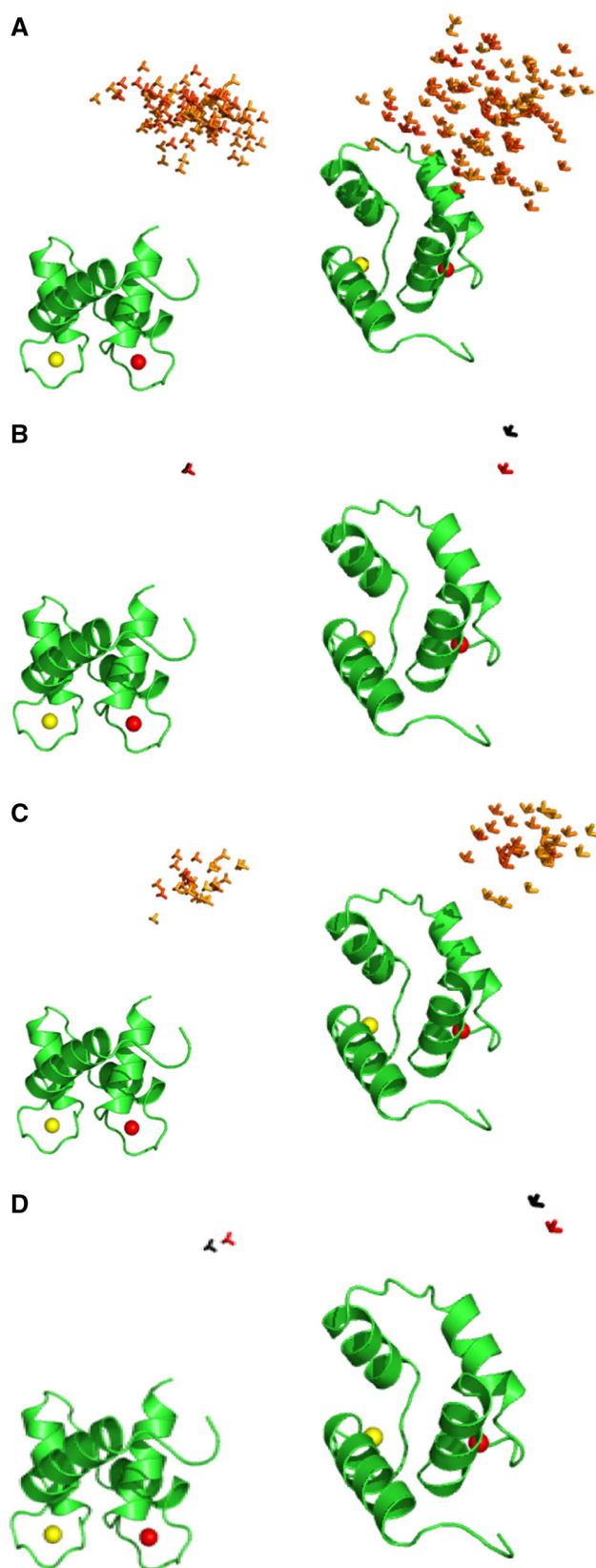
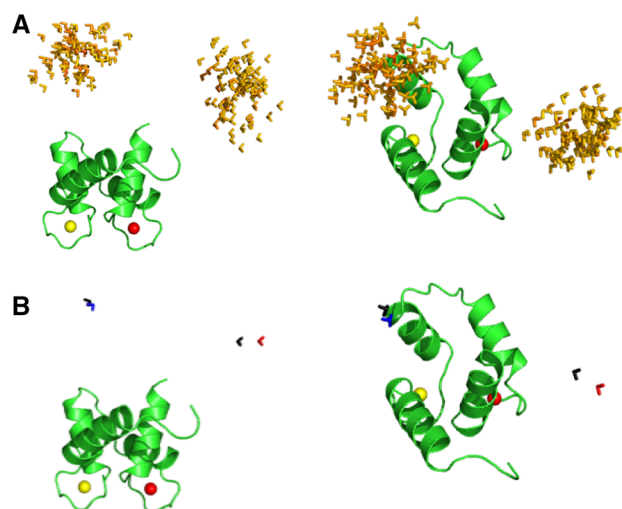◄**Fig. 6** MaxOcc results for Simulation 1. Each conformation is represented by a triad of Cartesian axes, centered at the center of mass of the C-terminal domain. Color code for **a**, **c**—according to the MaxOcc value (0.0—*blue*, 0.8—*red*), **a** The conformers with the highest MaxOcc recovered with the paramagnetic data (with error of 1 Hz for pRDC and 0.01 ppm for PCS), **b** The center of the distribution shown in *panel a* (*red*) versus the center of the simulated region (*black*), **c** The conformers with the highest MaxOcc recovered with dRDC (with error of 1 Hz), **d** The center of the distribution shown in *panel c* (*red*) versus the center of the simulated region (*black*). The results are shown from two different points of view in the *left* and *right panels*. All conformers are superimposed by the N-terminal domain, shown as a ribbon

Fig. 7 MaxOcc results for Simulation 2 with dRDC with error of 1 Hz. Each conformation is represented by a triad of Cartesian axes, centered at the center of mass of the C-terminal domain. **a** The conformers with the highest MaxOcc, color code: according to the MaxOcc value (0.0—*blue*, 0.6—*red*). **b** The center of the distribution shown in *panel a* (*red*) versus the center of the simulated region (*black*). All conformers are superimposed by the N-terminal domain, shown as a ribbon

reported in Table S1 (diagonal entries). All regions have similar MaxOR values (up to 0.60), not much higher than the largest MaxOcc values for the individual conformations. If however MaxOR values are calculated for pairs of regions (off-diagonal entries of Table S1), strong differences arise. All pairs yielding the highest MaxOR (0.90–1.00) are composed of regions at the opposite sides of the distribution of the highest MaxOcc conformers, whereas all pairs composed of the regions located on the same side of the distribution or more importantly containing a region in the middle, have significantly lower MaxOR (up to 0.63 and 0.78, respectively). This strongly suggests the occurrence of a two-site exchange model. The
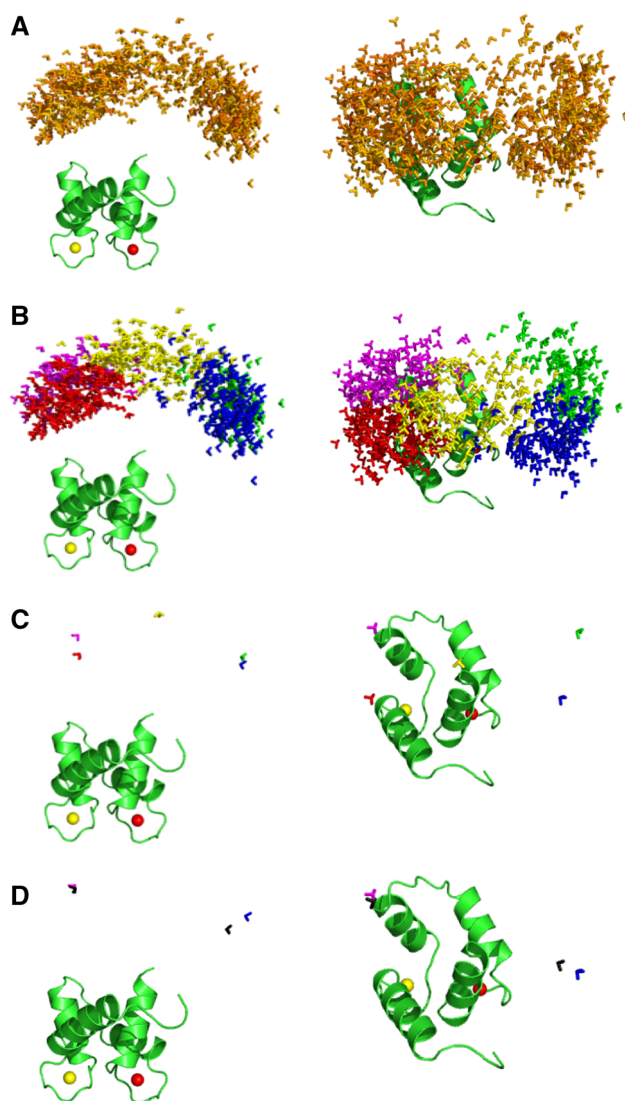
Fig. 8 MaxOcc/MaxOR results for Simulation 2 with paramagnetic data (1 Hz error for pRDC and 0.01 ppm error for PCS). Each conformation is represented by a triad of Cartesian axes, centered at the center of mass of the C-terminal domain. **a** The conformers with the highest MaxOcc, color code—according to the MaxOcc value (0.0—*blue*, 0.6—*red*). **b** The five clusters formed by the conformers in *panel a*. **c** The centers of the clusters. **d** The centers of clusters with the largest MaxOR versus the centers of the simulated regions (*black*). All conformers are superimposed by the N-terminal domain, shown as a ribbon

pair of regions with the highest MaxOR has their central conformations in nice agreement with the conformations in the center of the distributions in the synthetic ensemble, with an accuracy comparable to that obtained by SES (Table 4; Fig. 8d).

In simulation 3, for both the paramagnetic and diamagnetic data, the conformers recovered by MaxOcc form elongated regions comprising both the two centers and

conformers situated between them (Figs. S3A and S4A). MaxOR was thus applied in both cases. As in the previous simulation, no single region has MaxOR significantly higher than the others, but the analysis of pairs of regions indicated again the occurrence of a two-site exchange (Tables S2 and S3). The two central conformations of the synthetic ensemble were identified with good accuracy (Table 4; Figs. S3D and S4D) using both kinds of experimental restraints. Again, the results are robust, as increased errors did not largely affect the accuracy of the solutions.

The performed MaxOcc/MaxOR analysis, as it appears from Table 4 as a whole, confirms the conclusion from the SES results that paramagnetic and diamagnetic restraints are equally useful for the recovery of conformational ensembles.

Conclusions

In many experimental studies RDCs have been shown to be precious restraints for analyzing molecular conformational freedom (Montalvao et al. 2014; Ravera et al. 2014; Camilloni and Vendruscolo 2015; Torchia 2015). Here we compared paramagnetic and diamagnetic RDCs and found substantial differences in their information content in the case of multidomain proteins. We found that the information content of dRDC is larger than that of pRDC in terms of number of singular values, and this reflects the shape dependence of dRDC. However, since the internal alignment due to paramagnetism also gives rise to PCSs, the total informational content recovered in a paramagnetic experiment is at least on par with dRDCs.

We have performed several simulations to evaluate the capability of recovering the conformational variability of two-domain proteins by the use of two different approaches, SES and MaxOcc/MaxOR. The main states of the protein were recovered reasonably well for both paramagnetic and diamagnetic datasets, with both approaches (see Tables 1, 2, 3, 4 and also Table S4). Even for rather large experimental errors, we have found that both datasets still retain the ability of recovering the main conformational states, thus resulting appealing for the analysis of averaged experimental data possibly also in the case of large systems, where RDCs are affected by large errors. Of course, since the problem is underdetermined, a correct reconstruction of the main states may be unsuccessful for different rather unpredictable conformational distributions.

Such analysis suggests that pRDC + PCS provide a very promising alternative to dRDC data. It is important to note that this analysis does not include modeling error, which is harder to quantify. Therefore, our analysis does not capture the principal advantages of pRDC + PCS over

dRDC, in that it does not require assumption of a barrier model in order to predict the alignment. In addition, one has to consider that the interactions of the protein with the alignment medium might actually perturb the system, and that these interactions can occur on a timescale that is slower than the conformational averaging itself, so that the assumption that the measured dRDCs can be represented as a population-weighted average of the RDCs for the individual (rigid) conformers may fall short in representing the real physical picture.

Finally, the availability of a number of rigid lanthanide-binding tags nowadays may make the acquisition of three independent metal ion datasets more practical and safer than the acquisition and prediction of three independent alignment media. One current limitation of using metal ions is the low signal-to-noise ratio in pRDC and PCS data, which could potentially be improved with better technology and methodology.

References

Al-Hashimi HM, Valafar H, Terrell M, Zartler ER, Eidsness MK, Prestegard JH (2000) Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. J Magn Reson 143:402–406

Allegrozzi M, Bertini I, Janik MBL, Lee Y-M, Liu G, Luchinat C (2000) Lanthanide induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion. J Am Chem Soc 122:4154–4161

Andralojc W, Luchinat C, Parigi G, Ravera E (2014) Exploring regions of conformational space occupied by two-domain proteins. J Phys Chem B 118:10576–10587

Balayssac S, Bertini I, Bhaumik A, Lelli M, Luchinat C (2008) Paramagnetic shifts in solid-state NMR of proteins to elicit strucutral information. Proc Natl Acad Sci USA 105:17284–17289

Banci L, Bertini I, Bren KL, Cremonini MA, Gray HB, Luchinat C, Turano P (1996) The use of pseudocontact shifts to refine solution structures of paramagnetic metalloproteins: Met80Ala cyano-cytochrome c as an example. J Biol Inorg Chem 1:117–126

Banci L, Bertini I, Gori Savellini G, Romagnoli A, Turano P, Cremonini MA, Luchinat C, Gray HB (1997) Pseudocontact shifts as constraints for energy minimization and molecular dynamic calculations on solution structures of paramagnetic metalloproteins. Proteins Struct Funct Genet 29:68–76

Banci L, Bertini I, Huber JG, Luchinat C, Rosato A (1998) Partial orientation of oxidized and reduced cytochrome b₅ at high magnetic fields: magnetic susceptibility anisotropy contributions and consequences for protein solution structure determination. J Am Chem Soc 120:12903–12909

Barbato G, Ikura M, Kay LE, Pastor RW, Bax A (1992) Backbone dynamics of calmodulin studied by ¹⁵N relaxation using inverse detected two-dimensional NMR spectroscopy; the central helix is flexible. Biochemistry 31:5269–5278

Barthelmes K, Reynolds AM, Peisach E, Jonker HRA, DeNunzio NJ, Allen KN, Imperiali B, Schwalbe H (2011) Engineering encodable lanthanide-binding tags into loop regions of proteins. J Am Chem Soc 133:808–819

Bashir Q, Volkov AN, Ullmann GM, Ubbink M (2010) Visualization of the encounter ensemble of the transient electron transfer complex of cytochrome c and cytochrome c peroxidase. J Am Chem Soc 132:241–247

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

Berlin K, Castañeda CA, Schneidman-Dohovny D, Sali A, Nava-Tudela A, Fushman D (2013) Recovering a representative conformational ensemble from underdetermined macromolecular structural data. J Am Chem Soc 135:16595–16609

Bernadò P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129:5656–5664

Bertini I, Donaire A, Jiménez B, Luchinat C, Parigi G, Piccioli M, Poggi L (2001a) Paramagnetism-based versus classical constraints: an analysis of the solution structure of Ca Ln calbindin D₉ₖ. J Biomol NMR 21:85–98

Bertini I, Janik MBL, Lee Y-M, Luchinat C, Rosato A (2001b) Magnetic susceptibility tensor anisotropies for a lanthanide Ion series in a fixed protein matrix. J Am Chem Soc 123:4181–4188

Bertini I, Janik MBL, Liu G, Luchinat C, Rosato A (2001c) Solution structure calculations through self-orientation in a magnetic field of cerium (III) substituted calcium-binding protein. J Magn Reson 148:23–30

Bertini I, Longinetti M, Luchinat C, Parigi G, Sgheri L (2002a) Efficiency of paramagnetism-based constraints to determine the spatial arrangement of α-helical secondary structure elements. J Biomol NMR 22:123–136

Bertini I, Luchinat C, Parigi G (2002b) Magnetic susceptibility in paramagnetic NMR. Prog NMR Spectrosc 40:249–273

Bertini I, Gelis I, Katsaros N, Luchinat C, Provenzani A (2003) Tuning the affinity for lanthanides of calcium binding proteins. Biochemistry 42:8011–8021

Bertini I, Del Bianco C, Gelis I, Katsaros N, Luchinat C, Parigi G, Peana M, Provenzani A, Zoroddu MA (2004a) Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. Proc Natl Acad Sci USA 101:6841–6846

Bertini I, Fragai M, Lee Y-M, Luchinat C, Terni B (2004b) Paramagnetic metal ions in ligand screening: the Co^II matrix metalloproteinase 12. Angew Chem Int Ed 43:2254–2256

Bertini I, Luchinat C, Parigi G, Pierattelli R (2005) NMR of paramagnetic metalloproteins. ChemBioChem 6:1536–1549

Bertini I, Gupta YK, Luchinat C, Parigi G, Peana M, Sgheri L, Yuan J (2007) Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains. J Am Chem Soc 129:12786–12794

Bertini I, Luchinat C, Parigi G, Pierattelli R (2008) Perspectives in NMR of paramagnetic proteins. Dalton Trans 2008:3782–3790

Bertini I, Kursula P, Luchinat C, Parigi G, Vahokoski J, Willmans M, Yuan J (2009) Accurate solution structures of proteins from X-ray data and minimal set of NMR data: calmodulin peptide complexes as examples. J Am Chem Soc 131:5134–5144

Bertini I, Bhaumik A, De Paepe G, Griffin RG, Lelli M, Lewandowski JR, Luchinat C (2010a) High-resolution solid-state NMR structure of a 17.6 kDa protein. J Am Chem Soc 132:1032–1040

Bertini I, Giachetti A, Luchinat C, Parigi G, Petoukhov MV, Pierattelli R, Ravera E, Svergun DI (2010b) Conformational space of flexible biological macromolecules from average data. J Am Chem Soc 132:13553–13558

Bertini I, Calderone V, Cerofolini L, Fragai M, Geraldes CFGC, Hermann P, Luchinat C, Parigi G, Teixeira JMC (2012a) The catalytic domain of MMP-1 studied through tagged lanthanides. Dedicated to Prof. A.V. Xavier. FEBS Lett 586:557–567

Bertini I, Ferella L, Luchinat C, Parigi G, Petoukhov MV, Ravera E, Rosato A, Svergun DI (2012b) MaxOcc: a web portal for maximum occurence analysis. J Biomol NMR 53:271–280

Bertini I, Luchinat C, Nagulapalli M, Parigi G, Ravera E (2012c) Paramagnetic relaxation enhancements for the characterization of the conformational heterogeneity in two-domain proteins. Phys Chem Chem Phys 14:9149–9156

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

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

Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition (vol 5, pg 789, 2009). Nat Chem Biol 5:954

Bonvin AM, Brunger AT (1996) Do NOE distances contain enough information to assess the relative populations of multi-conformer structures? J Biomol NMR 7:72–76

Bothe JR, Nikolova EN, Eichhorn CD, Chugh J, Hansen AL, Al Hashimi HM (2011) Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. Nat Methods 8:919–931

Burgi R, Pitera J, Van Gunsteren WF (2001) Assessing the effect of conformational averaging on the measured values of observables. J Biomol NMR 19:305–320

Camilloni C, Vendruscolo M (2015) A tensor-free method for the structural and dynamical refinement of proteins using residual dipolar couplings. J Phys Chem B 119:653–661

Cerofolini L, Fields GB, Fragai M, Geraldes CFGC, Luchinat C, Parigi G, Ravera E, Svergun DI, Teixeira JMC (2013) Examination of matrix metalloproteinase-1 (MMP-1) in solution: a preference for the pre-collagenolysis state. J Biol Chem 288:30659–30671

Chen Y, Campbell SL, Dokholyan NV (2007) Deciphering protein dynamics from NMR data using explicit structure sampling and selection. Biophys J 93:2300–2306

Chou JJ, Li S, Klee CB, Bax A (2001) Solution structure of Ca^{2+} calmodulin reveals flexible hand-like properties of its domains. Nat Struct Biol 8:990–997

Choy W-Y, Forman-Kay JD (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. J Mol Biol 308:1011–1032

Chuang GY, Mehra-Chaudhary R, Ngan CH, Zerbe BS, Kozakov D, Vajda S, Beamer LJ (2010) Domain motion and interdomain hot spots in a multidomain enzyme. Protein Sci 19:1662–1672

Clore GM, Schwieters CD (2004) 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

Das Gupta S, Hu X, Keizers PHJ, Liu W-M, Luchinat C, Nagulapalli M, Overhand M, Parigi G, Sgheri L, Ubbink M (2011) Narrowing the conformational space sampled by two-domain

proteins with paramagnetic probes in both domains. J Biomol NMR 51:253–263

Diaz-Moreno I, Diaz-Quintana A, De la Rosa MA, Ubbink M (2005) Structure of the complex between plastocyanin and cytochrome f from the cyanobacterium nostoc Sp. PCC 7119 as determined by paramagnetic NMR. J Biol Chem 280:18908–18915

Fisher CK, Stultz CM (2011) Constructing ensembles for instrinsically disordered proteins. Curr Opin Struct Biol 21:426–431

Fisher CK, Huang A, Stultz CM (2010) Modeling intrinsically disordered proteins with bayesian statistics. J Am Chem Soc 132:14919–14927

Fragai M, Luchinat C, Parigi G (2006) "Four-dimensional" protein structures: examples from metalloproteins. Acc Chem Res 39:909–917

Gaponenko V, Sarma SP, Altieri AS, Horita DA, Li J, Byrd RA (2004) Improving the accuracy of NMR structures of large proteins using pseudocontact shifts as long/range restraints. J Biomol NMR 28:205–212

Gardner RJ, Longinetti M, Sgheri L (2005) Reconstruction of orientations of a moving protein domain from paramagnetic data. Inverse Probl 21:879–898

Gempf KL, Butler SJ, Funk AM, Parker D (2013) Direct and selective tagging of cysteine residues in peptides and proteins with 4-nitropyridyl lanthanide complexes. Chem Commun (Camb) 49:9104–9106

Gochin M, Roder H (1995a) Protein structure refinement based on paramagnetic NMR shifts: applications to wild-type and mutants forms of cytochrome c. Protein Sci 4:296–305

Gochin M, Roder H (1995b) Use of pseudocontact shifts as a structural constraint for macromolecules in solution. Bull Magn Reson 17:1–4

Guerry P, Salmon L, Mollica L, Ortega Roldan JL, Markwick P, van Nuland NA, McCammon JA, Blackledge M (2013) Mapping the population of protein conformational energy sub-states from NMR dipolar couplings. Angew Chem Int Ed Engl 52:3181–3185

Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol 5:1065–1074

Hass MAS, Keizers PHJ, Blok A, Hiruma Y, Ubbink M (2010) Validation of a lanthanide tag for the analysis of protein dynamics by paramagnetic NMR spectroscopy. J Am Chem Soc 132:9952–9953

Häussinger D, Huang J, Grzesiek S (2009) DOTA-M8: an extremely rigid, high-affinity lanthanide chelating tag for PCS NMR spectroscopy. J Am Chem Soc 131:14761–14767

Huang J, 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

Hulsker R, Baranova MV, Bullerjahn GS, Ubbink M (2008) Dynamics in the transient complex of plastocyanin-cytochrome f from Prochlorothrix hollandica. J Am Chem Soc 130:1985–1991

Iwahara J, Schwieters CD, Clore GM (2004) Ensemble approach for NMR structure refinement against H-1 paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. J Am Chem Soc 126:5879–5896

Jensen MR, Hansen DF, Ayna U, Dagil R, Hass MA, Christensen HE, Led JJ (2006) On the use of pseudocontact shifts in the structure determination of metalloproteins. Magn Reson Chem 44:294–301

John M, Otting G (2007) Strategies for measurements of pseudocontact shifts in protein NMR spectroscopy. ChemPhysChem 8:2309–2313

Jones E, Oliphant E, Peterson P et al (2001) SciPy: Open source scientific tools for Python

Keizers PHJ, Saragliadis A, Hiruma Y, Overhand M, Ubbink M (2008) Design, synthesis, and evaluation of a lanthanide chelating protein probe: CLaNP-5 yields predictable paramagnetic effects independent of environment. J Am Chem Soc 130:14802–14812

Kobashigawa Y, Saio T, Ushio M, Sekiguchi M, Yokochi M, Ogura K, Inagaki F (2012) Convenient method for resolving degeneracies due to symmetry of the magnetic susceptibility tensor and its application to pseudo contact shift-based protein-protein complex structure determination. J Biomol NMR 53:53–63

Korzhnev DM, Kay LE (2008) Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. Acc Chem Res 41:442–451

Kuffner JJ (2004) Effective sampling and distance metrics for 3D rigid body path planning. In: Proceedings IEEE international conference on Robotics and Automation (ICRA), vol 4, p 3993

Kukic P, Camilloni C, Cavalli A, Vendruscolo M (2014) Determination of the individual roles of the linker residues in the interdomain motions of calmodulin using NMR chemical shifts. J Mol Biol 426:1826–1838

Kurland RJ, McGarvey BR (1970) Isotropic NMR shifts in transition metal complexes: calculation of the Fermi contact and pseudo-contact terms. J Magn Reson 2:286–301

Lakomek NA, Walter KF, Fares C, Lange OF, de Groot BL, Grubmuller H, Bruschweiler 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 N-A, Farès 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

Latham MP, Hanson P, Brown DJ, Pardi A (2008) Comparison of alignment tensors generated for native tRNA(Val) using magnetic fields and liquid crystalline media. J Biomol NMR 40:83–94

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

Liu WM, Keizers PH, Hass MA, Blok A, Timmer M, Sarris AJ, Overhand M, Ubbink M (2012) A pH-sensitive, colorful, lanthanide-chelating paramagnetic NMR probe. J Am Chem Soc 134:17306–17313

Loh CT, Ozawa K, Tuck KL, Barlow N, Huber T, Otting G, Graham B (2013) Lanthanide tags for site-specific ligation to an unnatural amino acid and generation of pseudocontact shifts in proteins. Bioconjug Chem 24:260–268

Lohman JAB, Maclean C (1978) Alignment effects on high resolution NMR spectra induced by the magnetic field. Chem Phys 35:269–274

Longinetti M, Luchinat C, Parigi G, Sgheri L (2006) Efficient determination of the most favored orientations of protein domains from paramagnetic NMR data. Inverse Probl 22:1485–1502

Losonczi JA, Prestegard JH (1998) Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. J Biomol NMR 12:447–451

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

Luchinat C, Nagulapalli M, Parigi G, Sgheri L (2012a) Maximum occurence analysis of protein conformations for different distributions of paramagnetic metal ions within flexible two-domain proteins. J Magn Reson 215:85–93

Luchinat C, Parigi G, Ravera E, Rinaldelli M (2012b) Solid state NMR crystallography through paramagnetic restraints. J Am Chem Soc 134:5006–5009

Maltsev AS, Grishaev A, Roche J, Zasloff M, Bax A (2014) Improved cross validation of a static ubiquitin structure derived from high precision residual dipolar couplings measured in a drug-based liquid crystalline phase. J Am Chem Soc 136:3752–3755

Man B, Su XC, Liang H, Simonsen S, Huber T, Messerle BA, Otting G (2010) 3-Mercapto-2,6-pyridinedicarboxylic acid: a small lanthanide-binding tag for protein studies by NMR spectroscopy. Chem Eur J 16:3827–3832

Montalvao R, Camilloni C, De SA, Vendruscolo M (2014) New opportunities for tensor-free calculations of residual dipolar couplings for the study of protein dynamics. J Biomol NMR 58:233–238

Musiani F, Rossetti G, Capece L, Gerger TM, Micheletti C, Varani G, Carloni P (2014) Molecular dynamics simulations identify time scale of conformational changes responsible for conformational selection in molecular recognition of HIV-1 transactivation responsive RNA. J Am Chem Soc 136:15631–15637

Nesterov Y (2012) Efficiency of coordinate descent methods on huge-scale optimization problems. SIAM J Optim 22:341–362

Nodet L, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M (2009) Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. J Am Chem Soc 131:17908–17918

O'Leary DP (2009) Scientific computing with case studies. SIAM, Bangkok

Pickford AR, Campbell ID (2004) NMR studies of modular protein structures and their interactions. Chem Rev 104:3557–3566

Pintacuda G, John M, Su XC, Otting G (2007) NMR structure determination of protein-ligand complexes by lanthanide labeling. Acc Chem Res 40:206–212

Prestegard JH, Al-Hashimi HM, Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. Q Rev Biophys 33:371–424

Ramirez BE, Bax A (1998) Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. J Am Chem Soc 120:9106–9107

Ravera E, Salmon L, Fragai M, Parigi G, Al-Hashimi HM, Luchinat C (2014) Insights into domain-domain motions in proteins and RNA from solution NMR. Acc Chem Res 47:3118–3126

Rinnenthal J, Buck J, Ferner J, Wacker A, Furtig B, Schwalbe H (2011) Mapping the landscape of RNA dynamics with NMR spectroscopy. Acc Chem Res 44:1292–1301

Rodriguez-Castañeda F, Haberz P, Leonov A, Griesinger C (2006) Paramagnetic tagging of diamagnetic proteins for solution NMR. Magn Reson Chem 44:S10–S16

Russo L, Maestre-Martinez M, Wolff S, Becker S, Griesinger C (2013) Interdomain dynamics explored by paramagnetic NMR. J Am Chem Soc 135:17111–17120

Ryabov YE, Fushman D (2006) Analysis of interdomain dynamics in a two-domain protein using residual dipolar couplings together with 15 N relaxation data. Magn Reson Chem 44:S143–S151

Ryabov YE, Fushman D (2007) A model of Interdomain mobility in a multidomain protein. J Am Chem Soc 129:3315–3327

Saio T, Ogura K, Shimizu K, Yokochi M, Burke TR Jr, Inagaki F (2011) An NMR strategy for fragment-based ligand screening utilizing a paramagnetic lanthanide probe. J Biomol NMR 51:395–408

Schmitz C, Vernon R, Otting G, Baker D, Huber T (2012) Protein structure determination from pseudocontact shifts using ROSETTA. J Mol Biol 416:668–677

Schroeder R, Barta A, Semrad K (2004) Strategies for RNA folding and assembly. Nat Rev Mol Cell Biol 5:908–919

Sgheri L (2010a) Conformational freedom of proteins and the maximal probability of sets of orientations. Inverse Probl 26:035003-1–035003-19

Sgheri L (2010b) Joining RDC data from flexible protein domains. Inverse Probl 26:115021-1–115021-12

Sicheri F, Kuriyan J (1997) Structures of Src-family tyrosine kinases. Curr Opin Struct Biol 7:777–785

Simin M, Irausquin S, Cole CA, Valafar H (2014) Improvements to REDCRAFT: a software tool for simultaneous characterization of protein backbone structure and dynamics from residual dipolar couplings. J Biomol NMR 60:241–264

Stelzer AC, Frank AT, Bailor MH, Andricioaei I, Al Hashimi HM (2009) Constructing atomic-resolution RNA structural ensembles using MD and motionally decoupled NMR RDCs. Methods 49:167–173

Su XC, Otting G (2010) Paramagnetic labelling of proteins and oligonucleotides for NMR. J Biomol NMR 46:101–112

Su XC, Huber T, Dixon NE, Otting G (2006) Site-specific labelling of proteins with a rigid lanthanide-binding tag. ChemBioChem 7:1599–1604

Su XC, Man B, Beeren S, Liang H, Simonsen S, Schmitz C, Huber T, Messerle BA, Otting G (2008a) A dipicolinic acid tag for rigid lanthanide tagging of proteins and paramagnetic NMR spectroscopy. J Am Chem Soc 130:10486–10487

Su XC, McAndrew K, Huber T, Otting G (2008b) Lanthanide-binding peptides for NMR measurements of residual dipolar couplings and paramagnetic effects from multiple angles. J Am Chem Soc 130:1681–1687

Svergun DI, Petoukhov MV, Koch MHJ (2001) Determination of domain structure of proteins from X-ray solution scattering. Biophys J 80:2946–2953

Swarbrick JD, Ung P, Chhabra S, Graham B (2011a) An iminodiacetic acid based lanthanide binding tag for paramagnetic exchange NMR spectroscopy. Angew Chem Int Ed Engl 50:4403–4406

Swarbrick JD, Ung P, Su XC, Maleckis A, Chhabra S, Huber T, Otting G, Graham B (2011b) Engineering of a bis-chelator motif into a protein alpha-helix for rigid lanthanide binding and paramagnetic NMR spectroscopy. Chem Commun (Camb) 47:7368–7370

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

Tjandra N, Kuboniwa H, Ren H, Bax A (1995) Rotational dynamics of calcium-free calmodulin studied by ^{15}N-NMR relaxation measurements. Eur J Biochem 230:1014–1024

Tolman JR (2001) Dipolar couplings as a probe of molecular dynamics and structure in solution. Curr Opin Struct Biol 11:532–539

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 (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc Natl Acad Sci USA 92:9279–9283

Tonks NK (2006) Protein tyrosine phosphatases: from genes, to function, to disease. Nat Rev Mol Cell Biol 7:833–846

Torchia DA (2015) NMR studies of dynamic biomolecular conformational ensembles. Prog Nucl Magn Reson Spectrosc 84–85:14–32

Valafar H, Prestegard JH (2004) REDCAT: a residual dipolar coupling analysis tool. J Magn Reson 167:228–241

Wang H, Eberstadt M, Olejniczak ET, Meadows RP, Fesik SW (1998) A liquid crystalline medium for measuring residual dipolar couplings over a wide range of temperatures. J Biomol NMR 12:443–446

Wöhnert J, Franz KJ, Nitz M, Imperiali B, Schwalbe H (2003) Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings. J Am Chem Soc 125:13338–13339

Yagi H, Maleckis A, Otting G (2013a) A systematic study of labelling an alpha-helix in a protein with a lanthanide using IDA-SH or NTA-SH tags. J Biomol NMR 55:157–166

Yagi H, Pilla KB, Maleckis A, Graham B, Huber T, Otting G (2013b) Three-dimensional protein fold determination from backbone amide pseudocontact shifts generated by lanthanide tags at multiple sites. Structure 21:883–890

Zhang Y, Zuiderweg ER (2004) The 70-kDa heat shock protein chaperone nucleotide-binding domain in solution unveiled as a molecular machine that can reorient its functional subdomains. Proc Natl Acad Sci USA 101:10272–10277

Zhang Q, Throolin R, Pitt SW, Serganov A, Al Hashimi HM (2003) Probing motions between equivalent RNA domains using magnetic field induced residual dipolar couplings: accounting for correlations between motions and alignment. J Am Chem Soc 125:10530–10531

Zhuang T, Lee HS, Imperiali B, Prestegard JH (2008) Structure determination of a Galectin-3-carbohydrate complex using paramagnetism-based NMR constraints. Protein Sci 17:1220–1231

Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 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, Bax A (2001) Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. J Biomol NMR 20:365–377