Simultaneous Definition of High Resolution Protein Structure and Backbone Conformational Dynamics using NMR Residual Dipolar Couplings

Guillaume Bouvignies, Phineus R. L. Markwick, and Martin Blackledge* [a]

Introduction

Molecular motions, enabling changes in protein backbone or sidechain conformation, are thought to play a crucial role in both protein stability and function. [1–4] Despite the recognized importance of dynamics for biochemical activity, most approaches to protein structure determination, whether based on crystallographic or solution studies, propose three-dimensional atomic representations of a single configuration that takes little or no account of conformational fluctuation. Mo-
tional properties are routinely measured in solution, most com-
monly using nuclear magnetic resonance (NMR) spin relaxa-
tion, [5–8] where rapid fluctuations, up to the range of the char-
acteristic rotational correlation time of the molecule (around
10 ns for medium-size proteins in aqueous solution at room temperature) can be characterised, [9–11] or using rotating frame relaxation dispersion experiments that can detect confor-

mational exchange occurring on slower timescales. [12–13] However,
the ability to elucidate both structural and dynamic aspects
would provide direct access to the conformational space sam-
pred by the native protein, as well as leading to more accurate
average conformations. NMR is uniquely suited to this purpose,
with experimental techniques routinely probing time and en-
semble-averaged conformation-dependent observables. These
observables are generally used to extract a single conforma-
tion, but inherently encode, albeit in some potentially complex
way, detailed information on conformational dynamics occur-
ing on multiple timescales up to the millisecond range. These
slower timescales are of particular interest, firstly because they
are not probed routinely by spin relaxation, and secondly be-
cause functionally important biological processes, including
enzyme catalysis, [14] signal transduction, [15] ligand binding or al-
losteric regulation, [16] requiring collective motions involving
groups of atoms or amino acids, are expected to occur in this
time range. The quest to use NMR data to determine average
couplings, that become measurable under conditions of weak
alignment, provide sensitive probes of both molecular structure
and dynamics. Residual dipolar couplings are becoming increas-
ingly powerful for the study of proteins in solution. In this mini-
review we present their use for the simultaneous determination of
protein structure and dynamics.
than the difference in chemical shifts of the interchanging species $\Delta \delta$ (generally less than tens to hundreds of milliseconds) can be relatively easily identified, with significantly populated species giving rise to distinct chemical shifts. Conformational equilibria with exchange rates faster than $\Delta \delta$ on the other hand still give rise to a single peak representing all exchanging species. This case is therefore far more difficult to identify, and is commonly ignored in standard structure determination protocols. Such difficulties are compounded because cross-relaxation effects giving rise to the nOe are also strongly dependent on the nature and timescale of fast (ps–ns) motional fluctuations of the two interacting nuclei.\(^{[19]}\) These dynamic-averaging effects may however vary throughout the molecule, due to the different motional properties of diverse sites in the protein, that can result in different effective cross-relaxation rates for similar average distances. This possibility is also very difficult to allow for during structure refinement, as the necessary information cannot be known a priori and is in general too complex to be determined from the available data.

Measured cross relaxation rates are thus dependent on dynamic events occurring on very different timescales, whose effects on the measured interaction will depend strongly on the nature and the timescale of the motion, rather than a simple and unique dependence on distance. In combination with additional sources of uncertainty in measured relaxation rates, such as spin diffusion,\(^{[20]}\) it has long been accepted that precise determination of distances from standard nOe measured in proteins is unrealistic. This has led to the development of the successful adoption of pragmatic procedures based on the measurement of a very large number of imprecise distances. These are then used in combination with classical molecular mechanics approaches, to triangulate a unique molecular conformation in agreement with the experimental data.

Despite the prevalence, and obvious success of single copy structure determination approaches, the dual dependence of NMR data, coding for both structure and dynamics, has been recognised ever since the very early applications of NMR to structural biology. In most cases this occurred when measured experimental data, from small protein fragments or cyclic peptides, could not be explained in terms of a single conformation.\(^{[21,22]}\) In one such study conformational space was extensively sampled by creating conformers in agreement with subsets of experimentally determined distances and an equilibrium proposed in agreement with all measured nOe and rotator frame relaxation measurements.\(^{[23]}\) A number of methods were proposed to modify restrained molecular dynamic simulations to allow the evolution of different conformations that on average reproduce the measured data. Slightly differing techniques for averaging over time and ensemble were proposed, exploiting either multiple parallel trajectories (ensemble-averaged restrained molecular dynamics EARMD) over which the restraint must be fulfilled on average, or a restraint integral applied over the length of the trajectory (time-averaged restrained molecular dynamics TARMD).\(^{[24–27]}\) The efficiency of these approaches to describe the diverse degrees of motion present throughout the molecule depends to an extent on the chosen number of copies (EARMD) or the length of the memory function chosen before the restraint must be reapplied (TARMD). As mentioned above NMR structure calculation can be locally under-determined, even for single-copy approaches, implying that the extent of additional conformational space explored when using multiple copies or time-averaging is not easily defined. Recently motional amplitudes derived from spin-relaxation rates were used to define this ensemble conformational disorder, allowing the ensemble average to sample configurations that could give rise to observed fast motional characteristics of backbone and sidechains.\(^{[28]}\)

Combination of such constraints with nOe that may be averaged over slowly exchanging conformers faces the problem of mixing effective timescales for the different phenomena. The authors have therefore developed additional procedures to average over different populations for different experimental observables.\(^{[29]}\) More recently Nilges and co-workers have also used extensive conformational sampling, in combination with a Bayesian analysis, to reproduce experimental data from ensembles of structures selected on the basis of their ability to reproduce experimental data.\(^{[30]}\)

As discussed above however, the major difficulties of simultaneously extracting structural and dynamic information from proteins using NMR data are incurred because the principal restraint, the nOe, is susceptible to additional dynamic averaging phenomena that cannot easily be accounted for, even using time or ensemble averaging of restraints.

Herein, we describe how residual dipolar couplings (RDCs), measured under conditions of partial molecular alignment\(^{[31,32]}\) can be used to overcome many of these difficulties, and to provide access to direct, analytical determination of both structural and dynamic components of protein behaviour in solution.

### Residual Dipolar Couplings

The dipolar coupling between two spins 1/2 ($\gamma_i \gamma_j$) is described by the time and ensemble average of the dipolar Hamiltonian over all sampled orientations [Eq. (1)]:

$$D_i = \frac{\gamma_i \gamma_j \mu_0 h}{8 \pi s} \left( \langle P_3 \cos \theta(t) \rangle \right)$$

(1)

$rij$ is the distance between the two nuclei, $\gamma_i$ and $\gamma_j$ are the gyromagnetic ratios of the two spins, $h$ is Planck’s constant and $\mu_0$ the permittivity of free space. Note that the dipolar Hamiltonian depends on the orientation $\theta$ of the internuclear vector between the coupled spins, relative to the magnetic field, following a second order Legendre polynomial dependence ($P_3 \cos(\theta)$). Time and ensemble averaging of this function, denoted by the angular brackets, reduces the measured coupling to zero under the conditions of orientational averaging found in isotropic solution. In order to measure a residual coupling (RDC) in solution it is necessary to induce partial alignment, or order, in the sample. It has been shown over ten years ago\(^{[31]}\) that simple dissolution of a protein in a dilute liquid crystal solution of phospholipid bicelle, would allow the measurement of large (tens of Hertz) couplings, while retaining the high
quality spectra necessary for high resolution protein NMR. Very rapidly additional solvent systems were developed to provide partial alignment.\[33–39\] In the case of a macromolecule whose shape does not change significantly, the average in Equation (1) can be described as a convolution of the restricted motion of the solute molecules, defined by the average over all orientations of the molecule relative to the magnetic field, and the orientation of the interspin vector relative to the molecule (see Figure 1). The preferential orientational averaging of the molecule is commonly described in terms of an alignment tensor $A$ whose units are dimensionless, and whose trace is zero, reflecting its probabilistic nature.\[40] It is convenient to describe the measured couplings in terms of their orientation relative to this alignment tensor or principal axis system (PAS) common to the whole molecule. The orientation of the PAS or alignment tensor with respect to the coordinate frame of the molecule can in return be defined simply via a three-dimensional Euler rotation $R(\alpha, \beta, \gamma)$. One can describe the measured coupling in terms of $\theta, \phi$, the polar angles of the inter-spin vector in the eigenframe of the alignment tensor, with eigenvalues $A_{xx}$, $A_{yy}$ and $A_{zz}$ as in Equation (2):

$$D_{ij}(\theta, \phi) = -\frac{\gamma_i^2 \mu_B h}{8\pi^2 \eta_i^2} \left[ A_{xx} \cos^2 \theta + A_{yy} \sin^2 \theta \cos^2 \phi + A_{zz} \sin^2 \phi \right]$$

(2)

or as defined in Equation (3)

$$D_{ij}(\theta, \phi) = -\frac{\gamma_i^2 \mu_B h}{16\pi^2 \eta_i^2} \left[ A_{xx} (3 \cos^2 \theta - 1) + \frac{3}{2} A_y \sin^2 \theta \cos 2\phi \right]$$

(3)

where $A_x = A_{xx}/2$ is the axial component of the alignment tensor and $A_z = (1/3)(A_{xx} + A_{yy})$ is the rhombic component. The available orientations of an interaction vector for a single measured RDC in the presence of a known tensor are depicted in cartoon form in Figure 2 on the surface of a sphere.

Figure 1. Orientation of the internuclear vector in the principle axis system of the molecular alignment tensor. The angles are those described in Equation (3).

Figure 2. Cartoon representation of the dependence of measured dipolar couplings on the orientation of the internuclear vector. Dipolar coupling isocontours are shown as shaded bands: black/dark grey: positive coupling, white: intermediate and zero coupling, light grey: negative coupling. The axes represent the axes of the alignment tensor.

Protein Structure Determination Using RDCs Alone

A number of studies have shown the utility of combining RDCs with sparse distance and dihedral angle constraints derived from nOe, scalar coupling and chemical shift data to determine low resolution folds of small proteins.\[41–44\] Two approaches have also been developed using only RDCs as experimental constraints. One of these, Molecular fragment replacement,\[45\] searches databases for peptide strands present in proteins of known structure, and combines these segments to propose a complete protein fold. In parallel we attempted a de novo approach that places oriented peptide planes and tetrahedral junctions sequentially, as a function of the measured RDCs, in order to define the polypeptide fold.\[46\] The approach was named ‘meccano’ (molecular engineering calculations using coherent association of non-averaged orientations), and its feasibility was initially demonstrated using the dataset from ubiquitin in bicelles and charged bicelles measured by the Bax group,\[47,48\] and applied to the determination of the backbone conformation of the active site of a larger molecule, methionine sulfoxide reductase (27 kD).\[49\]

The initial step of the approach is to calibrate the alignment tensors, using no information concerning the molecular fold. This requires optimization of 7 parameters ($A_{xx}^1, A_{xx}^2, A_{xx}^3, A_{yy}^1, \alpha, \beta, \gamma$), where $\alpha, \beta, \gamma$ describe the orientation of $A^2$ with respect to $A^1$, taken to be diagonal in the calculation frame. Simultaneously, the orientation of each peptide plane is determined with respect to the calculation frame. This algorithm reliably finds the global minimum of the target function (Eq. (4)):

$$\chi^2 = \sum_i \left[ D_{ij}^{\text{exp}} - D_{ij}^{\text{calc}} \right]^2 / \sigma_i^2$$

(4)

over all measured couplings ($\sigma_i$ is the experimental uncertainty), requiring no a priori estimation of the alignment tensors. The second stage of the algorithm constructs the molecule. The orientation of any planar element has a twofold degeneracy with respect to the possible couplings in the presence of two different alignment media, introducing an orientational ambiguity between the correct alignment and the mirror image that must be solved for each planar element. RDCs measured in the tetrahedral junction ($A_{xx}^1 = -A_{xx}^2, A_{yy}^1 = -A_{yy}^2$) and the geometry of the junctions allow unambiguous positioning. The final Ubiquitin backbone conformation compares closely with the structure determined using an extensive nOe data set, in combination with RDCs, $^1J$ couplings and hydrogen...
bonding restraints (2.0 Å backbone rmsd, and 1.0 Å following refinement with a standard force field).

Protein Dynamics from RDCs

So far we have considered only the structural dependence of dipolar couplings and their application for determining static structures, but it is in terms of molecular dynamics that a second, equally powerful aspect of RDCs, is revealed. RDCs are averaged over all orientations of the magnetic dipolar interaction vector sampled up to a timescale defined by the inverse of the alignment-induced coupling, thus reporting on averages up to the millisecond range under conditions of partial molecular alignment.50

Expressing the dependence of the dipolar coupling on the vector orientation with respect to the alignment tensor as in Equation (3), we implicitly assume that the inter-spin vector is static with respect to the alignment tensor. In the presence of local internal motion the measured coupling is better represented by incorporating local conformational averaging over both time and ensemble [Eq. (5)]:

\[
\langle D_{ij}(\theta, \phi) \rangle = -\frac{\gamma_i \gamma_j \mu_h}{16 \pi^2 r_{ij}^3} \left[ A_1 (3 \cos^2 \theta - 1) + \frac{3}{2} A_2 (\sin^2 \theta \cos 2\phi) \right]
\]

(5)

The angular brackets indicate conformational averaging. This provides access to information that is potentially highly complementary to the dynamic parameters routinely extracted from spin relaxation measurements.51 Comparison of motional averaging on the two timescales provides information on dynamics in the nano- to millisecond range. This relevance is particularly evident for first order averaging of dipolar interactions whose rapid reorientation also dominates experimental spin relaxation rates (for example 1H couplings). The ability of RDCs to describe local conformational fluctuations over the nano to millisecond time range in proteins has been studied by a number of groups in recent years. The effects of intramolecular motion on the dynamic averaging of RDCs is represented in Figure 3. In the presence of different alignment tensors the averaging will sample dipolar couplings isocontours (shown as shaded bands in Figure 3) differently and give rise to differential averaging effects. A number of methods have been developed that attempt to extract the extent and shape of the motional envelope of internuclear vectors from dipolar couplings measured in differently aligning media.

Prestegard and co-workers interpreted local motions in terms of local alignment characteristics, and expressed these as a site-specific generalized degree of order (GDO),52 while Griesinger and co-workers used a very large number of datasets measured on Ubiquitin53–56 to determine the shape and size of the orientational averaging envelope for each N–H vector in the protein. Tolman has independently proposed and applied related approaches.57,58 In an alternative approach, Bax and co-workers have attempted to define the limits of possible local dynamic amplitudes in protein GB3 by using refined static models, and comparing the ability of these models to describe the experimental data.59 Finally, with reference to earlier discussion, Clore et al. have used ensemble averaging of RDCs to describe the conformational space available to both Ubiquitin and protein GB3.60

We have explored the possibility of using specific geometric models to describe local motional averaging of RDCs. Initial studies used a one-dimensional gaussian axial fluctuation (GAF) model for peptide plane reorientation about the C\(^\alpha\)–C\(^\beta\)–C\(^\gamma\) axis, identifying a common anisotropic component of protein backbone dynamics from \(^{15}\)N–\(^1\)H RDCs.61–63 This simple approach demonstrated that statistically significant improvements could be made to the accuracy of the description of the overall molecular alignment tensor by taking into account local motions, even in the absence of site-specific details. In the light of vigorous debate concerning the nature and extent of slow motions present in soluble proteins, and the ability of RDCs to describe these dynamics, we undertook a detailed study of the presence of slow motions in protein GB3, interpreting an extensive set of RDCs64,65 measured in multiple differently aligning media, in terms of the three dimensional gaussian axial fluctuation model (3D-GAF, see Figure 4). This general model of peptide plane dynamics allows for stochastic motions around three orthogonal axes attached to the peptide plane,65 and our interpretation assumed a fixed time- and ensemble-averaged model of the average structure.

The study delivered a site-specific motional description of each peptide plane in the protein, and provided a quantitative estimate of the nature and extent of dynamics present on the protein backbone.66 We identified a heterogeneous distribution of slower motions in the protein in comparison to \(^{15}\)N spin relaxation data,67 with local motions in some regions of the protein that are quantitatively the same as those detected using spin relaxation, for example in the α-helix and some sur-
face loops and turns. We can therefore detect no additional (ns–ms) slow motions in these regions. In the β-sheet, and one of the surface loops, however, slower motions are observed, in particular in the region where the protein interacts with its physiological partner (β-strand II). The presence of these dynamic modes is verified using extensive cross validation of data that are not used in the analysis, and the dependence on the structural model is tested against two crystal conformations and an RDC refined structure, all of which gave similar motional distributions.

Analysis of trans-hydrogen bond scalar couplings in terms of these local dynamic amplitudes and directions also found strong evidence that the motion was correlated and that the collective motion transmitted across the β-sheet was propagated via the inter-strand hydrogen bonds. Although this kind of transmission of dynamics has been proposed, it has never been observed by other experimental methods, and is computationally challenging to simulate. The existence of these slow motional modes extending across the entire β-sheet carries clear implications for understanding the mechanisms of long-range signal propagation in proteins. In the case of protein G, these findings illustrate how the protein harnesses thermal motions via specific dynamic networks to enable molecular function at the interaction site (see Figure 5).

This parameterisation of molecular motion in terms of a simple dynamic model led us to explore the possibility of using extended sampling molecular dynamics based approaches to develop a more explicit molecular description of the slower timescale motions present in GB3. Accelerated molecular dynamics (AMD) simulations were shown to accurately reproduce the amplitude and distribution of slow motional modes characterized using RDCs and the 3D GAF model. In agreement with experiment, larger amplitude slower motions were localized in the β-strand/loop-motif spanning residues 14–24 and in loop 42–44. Principal component analysis showed these fluctuations dominating the lowest energy eigenmode, substantiating the existence of a correlated motion traversing the β-sheet that culminates in maximum excursions at the active site of the molecule. Interestingly fast dynamics were simulated using extensive standard MD simulations and compared to order parameters extracted from 15N relaxation. Sixty 2 ns fully-solvated MD simulations exploring the different conformational sub-states sampled from AMD resulted in significantly better reproduction of order parameters compared to the same number of simulations starting from the relaxed crystal structure, illustrating the inherent dependence of protein dynamics on local conformational topology. The results provide rare insight into the complex hierarchy of dynamics present in proteins and allow us to develop a picture of the conformational landscape native to the protein. Recent results from Showalter and Brüschweiler have also demonstrated that free MD simulations using the most recent force fields provide very close reproduction of experimental RDCs from Ubiquitin, substantiating the conclusions drawn from our study.

We have also used standard and accelerated molecular dynamics simulations to determine the accuracy of 3D GAF-based approaches to characterising the nature and extent of local molecular motions. The results demonstrate the robustness of the 3D GAF analysis even in the presence of large-scale motions, and illustrate the remarkably quantitative nature of extracted amplitudes. These observations suggest that the approach can generally be employed for the study of functionally interesting biomolecular motions.

**Simultaneous Determination of Protein Structure and Dynamics from RDCs**

It appears therefore that RDCs can be used both to determine high resolution protein structure, and to accurately describe the nature and extent of local molecular motions, clearly illustrating the dual dependence of these parameters, and underlining their remarkable potential for studying biomolecular behaviour in solution. In comparison to NOe, experimental RDCs are sensitive to relatively few sources of uncertainty, and as we have seen, are averaged over all conformations sampled up to the millisecond range, simplifying their interpretation in terms of both structure and dynamics. This means that dynamic effects can in principle be analytically incorporated into the con-
formational definition of the coupling, suggesting these data as optimal candidates for simultaneous determination of protein structure and dynamics.

In a recent study we therefore attempted to use peptide plane RDCs alone to simultaneously determine the backbone structure and conformational dynamics of protein GB3. A version of the meccano approach was developed, using analytical descriptions of the structural and dynamic dependence of RDCs. In particular we exploit the dependence of measured RDCs on the orientation and a single anisotropic motion about each of the 3DGAF axes [Eq. (6)]:

\[
\langle D_{ij}\rangle_{3DGAF} = -\frac{\mu_0}{8\pi^2} \frac{\gamma^2 I_h^J}{I_0^J} \times \\
\left\{ \begin{array}{l}
\frac{1}{2} D_{ij} \left[ (3 \cos^2 \beta - 1) + 3s_3 \sin 2\beta \cos \alpha + 3s_2 \sin^2 \beta \cos 2\alpha \right] + \\
\frac{3}{8} D_{ij} \left[ s_1 \sin^2 \beta \cos 2\gamma - 2s_5 \sin \beta \left( \cos(2\alpha + 2\gamma) \cos \frac{\beta}{2} - \cos(2\alpha - 2\gamma) \sin \frac{\beta}{2} \right) \right] + \\
+2s_5 \left( \cos(2\alpha + 2\gamma) \cos \frac{\beta}{2} + \cos(2\alpha - 2\gamma) \sin \frac{\beta}{2} \right) \end{array} \right\} 
\]

where \( s_1 = 2(3 \cos^2 \theta - 1) \), \( s_2 = 2 \sin \theta \exp(-\sigma^3/2) \), \( \sigma = 2 \sin^3 \theta \exp(-2\sigma^3) \) and \( r_i \) is the internuclear distance, \( \sigma \) the motional amplitude, \( R(\alpha, \beta, \gamma) \) the Euler transformation imposing the alignment tensor principal axis system into the local peptide plane frame defined with its \( z \) axis along the rotational axis and its \( x \) axis so that the internuclear vector lies in the \( x \)–\( z \) plane. \( \theta \) is the azimuthal angle between the rotational axis and the internuclear vector.

This relationship was used to determine both the average peptide plane orientation and the amplitude of the major mode of dynamic reorientation for each peptide plane in the protein. Three coplanar RDCs measured in five alignment media are used, comprising 750 \(^{15}N\)–\(^{1}H\), \(^{13}C\)–\(^{1}H\) and \(^{15}N\)–\(^{13}C\)–\(^{1}H\) RDCs. Only peptide plane RDCs are used in this study, as these are the structural units for which we have an analytical model of conformational dynamics. The only a priori knowledge of the protein structure used here is an assumed common, perfectly flat, peptide plane conformation derived from a database of ultra-high resolution X-ray crystallographic structures.

As in the case of the static-mecanno approach, the first step in the dynamic-mecanno protocol is to determine the components \( D_{ij}, D_{ij}, \theta, \phi, \psi \) of all five tensors. This entails the determination of the average orientation of each plane \( (\alpha, \beta, \gamma) \), a parameter \( \sigma \) accounting for dynamic fluctuation of each plane and the 22 parameters defining the tensors. A total of 238 parameters are therefore optimized in this step. Dynamic reorientation was incorporated using the best-fitting model for each peptide plane from 3 orthogonal one dimensional Gaussian axial fluctuations (1DGAF) or a common scaling factor for all RDCs in the plane. In the interests of comparison molecular alignment tensors were also determined using the static model, in Equation (3). Some motion may in this case be absorbed into the fitted average tensor components \( D_{ij}^{av} \) and \( D_{ij}^{av} \).

The backbone conformation is sequentially built by positioning the average conformations of the peptide planes of fixed internal geometry to best reproduce the experimental data by minimizing Equation (4). Intervening tetrahedral junction geometries were also weakly imposed. Optimisation of the plane orientation is accompanied by selection of one of the four dynamic modes described above (amplitudes \( \sigma_{\alpha}, \sigma_{\beta}, \sigma_{\gamma} \) of the GAF motions, or an order parameter \( S \), of an axially symmetric motion) on the basis of the target function for this plane. Once the structure has been determined on the basis of this algorithm a complete 3DGAF analysis of the dynamic disorder present along the chain was applied.

The advantage of such a study is the ability to directly compare the quality of optimally performed static and dynamic descriptions of the protein in solution, without resorting to the use of molecular mechanical force fields and in the near absence of spurious contributions to the experimental data. In this respect the results are remarkable. The dynamic-mecanno structure (pdb accession code 2nmq) directly resulting from this procedure is shown in Figure 6 in comparison to the 1.1 Å crystal structure (1igd: backbone rmsd 0.55 Å) and to the crystal structure refined with respect to the same RDCs (1p7e: backbone rmsd 0.34 Å). The dynamic-mecanno structure is clearly very close to these conformations. The main differences between the dynamic-mecanno structure and the X-ray structure occur in the loop region 14–20, where our previous 3DGAF study revealed the presence of slow dynamics. The static-mecanno structure also finds the global fold correctly, but is significantly further from 1igd and 1p7e (rmsd of 1.15 Å and 1.10 Å respectively) than the dynamic model. Of

Figure 6. Comparison of dynamic-mecanno and static-mecanno structures (red) with known conformations (blue) [left: 1igd the crystal structure, right: 1p7e the RDC-refined crystal structure]. Static meccano gives structures whose rmsds are 1.15 Å and 1.10 Å respectively, while dynamic meccano gives structures that are 0.55 Å and 0.34 Å respectively.
course this does not immediately tell us which of these models is more accurate, as both the crystal and the RDC-refined crystal structures may actually be different to the true solution average conformation. Structure determination using only peptide plane RDCs however can address this issue and tell us whether the dynamic-meccano description is actually better than the static-meccano structure. Using classical cross validation approaches, five calculations were performed for both static and dynamic-meccano calculations, with all data from one of the media removed from the analysis in each case. Predicted values were compared to experimental values for both static and dynamic cases. The dynamic-meccano structure was found to reproduce experimental data significantly better than the static-meccano structure (reduced χ² of 0.8 compared to 1.7), even for the most linearly independent alignment medium (χ²_red of 0.6 compared to 2.6), with largest discrepancies in the static case corresponding to the most dynamic sites in the protein. Interestingly the dynamic-meccano average conformation is further validated by 250 C−H RDCs that were not used in the calculation but are reproduced significantly better from this structure than by 1igd (reduced χ² of 7.8 compared to 14.4) (Figure 7). Although the C−H RDCs are interpreted in a static way in both cases, the average dynamic meccano structure still provides an improved average conformation as measured by the orientation of these vectors present in the tetrahedral C° centres.

Anisotropic dynamic parameters extracted from the 3DGAF analysis of the dynamic-meccano structure compare closely to values determined with respect to 1p7e, and to those predicted using the recent accelerated molecular dynamics study of protein G (Figure 8). Importantly motion about the γ-axis (C−C° direction) and the effective order parameter S°_NH RDC are similarly distributed along the chain and of similar amplitude. Alternation of anisotropic motions in the β-strand, indicative of a complex correlated motion across the β-sheet, is again observed. The S°_NH RDC values are in general closely reproduced between analyses of the dynamic-meccano and 1p7e structures, and are similarly distributed to those from a recent study using EARMD of the same RDCs. The distribution and amplitude of dynamics determined here are therefore very similar to those recently evaluated using the 3DGAF model applied to both 1p7e and 1igd.

This study also allowed us to address the remaining questions concerning structural noise for interpreting RDCs using an existing conformation. Dynamic-meccano constructs the backbone conformation ab initio, so that this source of error is not a concern. The similarity of these results with 1p7e, where peptide units are not all flat (3.1±2.0° deviation) also implies that this effect does not strongly influence extracted dynamic amplitudes. Recent comparisons of data simulated from accelerated molecular dynamics trajectories demonstrates that the dynamic-meccano protocol closely reproduces the average coordinates of the trajectory even in the case of broad conformational sampling. This result substantiates conclusions of our previous study; including the observation that slow motions are not uniformly distributed through the protein backbone, but are present only in certain regions, and affirms the independence of the previous results on the structural model. This important result also demonstrates the accuracy and robustness of these GAF-based approaches, that appear to allow quantitative estimates of the extent of slow motions in proteins at ambient temperature. This quantitative analysis relies on a careful treatment of the level of molecular alignment with respect to which the dynamic amplitudes are estimated.

It is interesting to piece together the information derived from these related studies of protein GB3 in order to better understand the conformational landscape of this small model protein. The evidence from RDC analysis and molecular dynamics simulation both suggest that the protein lies in a
steep-sided conformational energy potential, whose stippled base is relatively shallow and apparently harmonic, and comprises multiple local minima. The average solution structure determined using only RDCs, and accounting for significant levels of dynamic sampling as described above turns out to be very close to the single-copy RDC refined structure determined by Ulmer et al. This may not be a general result, but in this case supports the model of a simple harmonic potential well. The crystal conformation on the other hand is slightly different and probably lies in distinct conformational minimum on the potential-energy landscape.

The apparent resolution of the dynamic-meccano structure as measured using cross validation is remarkable in view of the simplicity of the approach. These results, in addition to the steep-sided conformational energy minimum on Ubiquitin [70] and protein GB3 [68], RDCs can be used to guide combination of conformational analysis of RDCs with the results of slow motions probed by native proteins and probably lies in distinct conformational minimum on the potential-energy landscape.

Conclusions and Outlook

In summary we have demonstrated the capacity of RDCs to simultaneously quantify slow motions probed by native proteins and their average conformations to high resolution. One of the most promising aspects of this work may result from the combination of conformational analysis of RDCs with the results of molecular dynamics simulation as shown by recent simulation on Ubiquitin [70] and protein GB3 [68]. RDCs can be used to guide molecular dynamics sampling methods, and provide a complete description of conformational behaviour of proteins in solution.

Acknowledgements

G.B. and P.R.L. M. receive grants from the CEA. This work was supported by the EU through EU-NMR JRA3 and French Research Ministry through ANR NT05-4 42781. The authors would like to thank the Centre de Calcul Recherche et Technologie (CCRT) (CEA, Bruyères-le-Châtel).

Keywords: conformation analysis · NMR spectroscopy · proteins · residual dipolar couplings · structure elucidation


Received: May 20, 2007
Published online on [ ] 2007
Simultaneous Definition of High Resolution Protein Structure and Backbone Conformational Dynamics using NMR Residual Dipolar Couplings

The whole story: In NMR data the nuclear overhauser effect is susceptible to additional dynamic averaging phenomena. Residual dipolar couplings measured under conditions of partial molecular alignment overcome these difficulties and provide access to direct, analytical determination of both structural and dynamics components of protein behaviour (see figure) in solution.