Protein Dynamics

DOI: 10.1002/anie.200900476

Protein Conformational Flexibility from Structure-Free Analysis of NMR Dipolar Couplings: Quantitative and Absolute Determination of Backbone Motion in Ubiquitin**

*Loïc Salmon, Guillaume Bouvignies, Phineus Markwick, Nils Lakomek, Scott Showalter, Da-Wei Li, Korvin Walter, Christian Griesinger, Rafael Brüschweiler, and Martin Blackledge**

Molecular dynamics play an essential role in controlling the biological activity of proteins. NMR residual dipolar couplings (RDCs)[1,2] are uniquely sensitive to conformational detail and thus offer a very attractive approach to the characterization of protein dynamics on all time scales up to the millisecond. The simple averaging properties of RDCs make them amenable to rigorous interpretation in terms of protein structure and dynamics. Assumptions made when analyzing protein dynamics from RDCs, as well as the robustness of the resulting dynamic description, can be largely substantiated through self-consistency checks. The recognition of these clear advantages has led to the development of several approaches to the characterization of protein-backbone motions from RDCs, including analytical deconvolution of the amplitudes and anisotropies of bond vectors or structural motifs,[3–8] ensemble-averaging by restrained-molecular-dynamics simulation,[9–11] and direct comparison to unrestrained molecular dynamics (MD).[12,13] In previous studies, however, only relative motional amplitudes could be determined directly from RDCs, because internal motional amplitudes and alignment strength cannot be separated easily. To derive absolute motional amplitudes from RDCs, NMR-relaxation-derived S^2_{Rel} order parameters have been used as upper limits of the corresponding RDC-derived order parameters S^2_{RDC}.[4,10,14]

Herein we describe a robust procedure for the quantitative and absolute determination of protein-backbone motions from RDCs that requires no scaling to an external reference, such as Lipari–Szabo order parameters. We have developed a novel, structure-free approach for the determination of the average orientation of each independent peptide plane in the protein and the associated local conformational dynamics about this mean. Motion is described by using the GAF (Gaussian axial fluctuation)[15–17] model of peptide-plane reorientation, whereby the amplitude, direction, and distribution of peptide-plane motions are analytically described. All alignment-tensor elements are determined simultaneously, which enables the quantitative assessment of the distribution of dynamic amplitudes. The approach was used to describe the conformational dynamics of the protein ubiquitin from experimental RDCs,[10,14,18–20] and the results are compared to motional modes extracted from a long (400 ns) MD simulation.

RDCs subject to conformational averaging in a folded protein can be expressed in terms of orientational bond-vector averages relative to a common molecular alignment frame:

$$D_i^j = -2D_a^j\sqrt{\frac{4\pi}{5}}\left\{\langle Y_0^2(\theta_i,\phi_i)\rangle + \sqrt{\frac{3}{8}}R\left(\langle Y_2^2(\theta_i,\phi_i)\rangle + \langle Y_{-2}^2(\theta_i,\phi_i)\rangle\right)\right\} \quad (1)$$

in which $D_a^j = -A_a^j\left(\mu_0\gamma_k\gamma_l h/16\pi^3 r_i^3\right)$ and R is the rhombicity. When analyzing the spherical harmonic averages $\langle Y_m^2(\theta_i,\phi_i)\rangle$ from RDCs, it is important to accurately determine the alignment tensors (j), and in particular the amplitude term for each tensor A_a^j. If tensors are determined without considering motion, average components of anisotropic and isotropic motional modes are absorbed into A_a^j so that the effective molecular alignment appears lower.[21] As a result, dynamic amplitudes extracted by any of the proposed approaches are reduced. Additional uncertainty may result from the coordinate sets used to estimate alignment tensors.

In this study, we addressed these issues by applying a structure-free GAF approach (SF GAF) to the elucidation of local motions from RDCs. Although in principal the method requires as few as three independent alignment media, in this case $^1D_{NH}$ couplings from 24 alignment media, $^1D_{C'N}$ and $^2D_{C'NH}$ couplings from five alignment media, and $^1D_{C'C_a}$ couplings from two alignment media were used. These data sets were chosen previously from a larger data set on the basis of self-consistency in terms of structure and dynamics.[22,14]

[*] L. Salmon,[+] Dr. G. Bouvignies,[+] Dr. P. Markwick, Dr. M. Blackledge
Protein Dynamics and Flexibility
Institute de Biologie Structurale Jean-Pierre Ebel
CNRS-CEA-UJF UMR 5075
41 rue Jules Horowitz, 38027 Grenoble Cedex (France)
Fax: (+33) 438-789-554
E-mail: martin.blackledge@ibs.fr
Dr. S. Showalter, Dr. D. W. Li, Prof. R. Brüschweiler
Chemical Sciences Laboratory
Department of Chemistry and Biochemistry and
National High Magnetic Field Laboratory (NHMFL)
Florida State University, Tallahassee, FL 32306 (USA)
Dr. N. Lakomek, K. Walter, Prof. C. Griesinger
Department of NMR-Based Structural Biology
Max Planck Institute for Biophysical Chemistry
Am Fassberg 11, 37077 Goettingen (Germany)
[+] These authors contributed equally to this work.

[**] This research was supported by the EU through EU-NMR JRA3 and by the French Research Ministry through ANR-07-PCVI-0013

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200900476.

The following function is minimized for each peptide plane for which a sufficient number of RDCs (>15) are available:

$$\chi^2\left[\{\theta,\phi,\psi\}_i,\{A\}_j,\{S,\sigma_\alpha,\sigma_\beta,\sigma_\gamma\}_i\right]=\sum_{ij}\left(D_{ij}^{\exp}-D_{ij}^{\mathrm{calc}}\right)^2/\delta_{ij}^2 \qquad (2)$$

in which the angles (θ,ϕ,ψ) describe the mean orientation of plane i, $\boldsymbol{A}$ represents the alignment tensor j, σ_α, σ_β ,σ_γ, and S are the amplitudes of the GAF or axially symmetric motions, and δ_{ij} is the estimated weighting of each RDC dataset, as determined by using a robust scaling estimator.[23] Because every peptide plane is treated separately from the others, this approach is "structure-free"; that is, the 3D protein fold is neither required nor constructed, in contrast to previous GAF-based applications.[6,7] The internal geometry of each peptide plane is defined by average heavy-atom coordinates extracted from ultra-high-resolution protein crystal structures.[6] The position of the H^N atom was optimized previously from extensive RDCs measured in protein G, with an optimal average $N{-}H^N$ distance of 1.02 Å,[21] consistent with ^{15}N-relaxation analysis. A more recent analysis suggests a distance of around 1.024 Å.[24] The use of this distance results in essentially identical results (see the Supporting Information). This procedure determines the optimal level of alignment relevant to the dominant motional mode. The result is then refined by cross-validation of "free datasets" by using the full 3D GAF description, which results in slightly higher tensors, probably owing to small-amplitude axially symmetric components that are otherwise absorbed into all A_a^j values. The procedure was applied to RDCs simulated from an MD trajectory of protein G;[25] four known alignment tensors were used, and the procedure was shown to be both valid and accurate to around 1% (see the Supporting Information).

By using the tensors defined in this way, 3D GAF analyses were applied to determine either I $\{\sigma_\gamma\}$, II $\{\sigma_\beta,\sigma_\gamma\}$, or III $\{\sigma_\alpha,\sigma_\beta,\sigma_\gamma\}$. For models I and II, values of $<\sigma_\alpha>=4°$ (I, II) and $<\sigma_\beta>=6°$ (I) were determined by optimizing averages from throughout the molecule. To avoid overfitting, models II and III were only invoked if the increased complexity of the model was statistically justified (on the basis of standard F-tests or AIC);[26] otherwise the average value was retained.

A comparison of the order parameters determined for the $N{-}H^N$ bond vectors by 3D GAF analysis (S^2_{RDC}) with those extracted from the trajectories of a long-timescale (400 ns) MD simulation (S^2_{MD})[12] revealed a similar profile and similar amplitude of motion (Figure 1 A). On average, the nature of the dynamics is very similar, as revealed by a comparison of motional amplitudes about the γ ($^{\alpha}C{-}^{\alpha}C$) axis (Figure 1 B). Only a few values that are close to zero, and are therefore expected to have a large error, show deviations.[25] S^2_{RDC} values were also compared to spin-relaxation order parameters S^2_{Rel}[27] (Figure 1 A). Within experimental uncertainty, three S^2_{RDC} values were found to be significantly higher than S^2_{Rel} (assuming an uncertainty of 0.03 for S^2_{Rel}): a nonphysical situation that may be partly caused by uncertainties in relaxation-data analysis. In previous analyses of the $N{-}H^N$ RDCs by self-consistent RDC-based model-free analysis (SCRM), alignment tensors were determined by using a static model, values were extracted for $\langle Y_m^2(\theta_i,\phi_i)\rangle$, and a

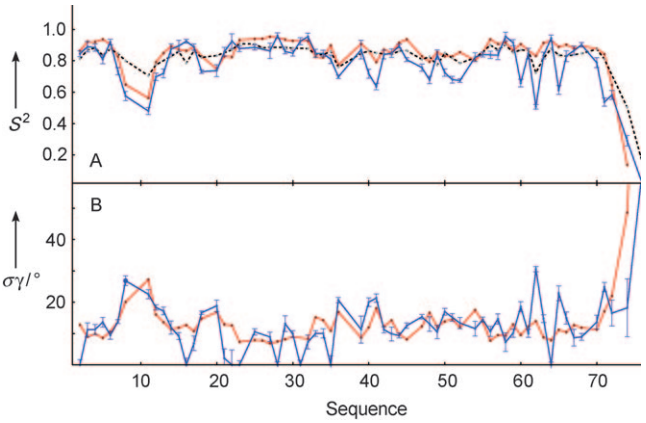

Figure 1. A) $N{-}H^N$ order parameters determined from SF GAF analysis of RDCs for ubiquitin (S^2_{RDC}: blue), 400 ns MD simulation (S^2_{MD}: red), and ^{15}N spin relaxation (S^2_{Rel}: dashed line). B: Amplitude of the γ component of 3D GAF motion (σ_γ) as determined from SF GAF RDC analysis (blue) and 400 ns MD simulation (red). Error bars are derived from noise-based Monte Carlo simulation.

scaling of S^2_{RDC} was applied to fulfill $S^2_{RDC} \le S^2_{Rel}$.[28,14] Applied to the relaxation values shown in Figure 1, this procedure revealed similar order parameters (Figure 2); it reproduced the details of the sequence-dependent dynamics found by using the 3D GAF approach.

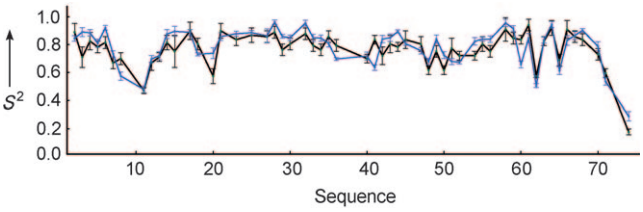

Figure 2. Comparison of $N{-}H^N$ order parameters determined from SF GAF (blue) and SCRM (black) values. The SCRM values were scaled to be equal to or lower than the relaxation data set.[14]

It was instructive to investigate the dependence of these findings on the magnitude of the alignment tensors (Figure 3). An overall scaling factor applied to all tensors was varied over a range $K_A=0.9$–1.1 of the determined A_a^j values. For each point, the local dynamic analysis was repeated with the scaled tensors. The resulting χ^2 value over the entire molecule was calculated for axially symmetric motion (S; model 1), for the best-fitting 1D GAF or axially symmetric motion (1D GAF/S) for each individual plane (model 2), and by 3D GAF analysis (model 3). Model 2 shows a minimum due to anisotropic motion of peptide planes, which dominates the detectable motion, and provides a systematically lower χ^2 value than that obtained with heterogeneously distributed axially symmetric models (if isotropic motion were more appropriate, the curve would be the same as for model 1). The 3D GAF analysis does not show a minimum, as this model can accommodate axially symmetric components while retaining local anisotropy. However, cross-validation of "free datasets" by using the full 3D GAF description does show a minimum,

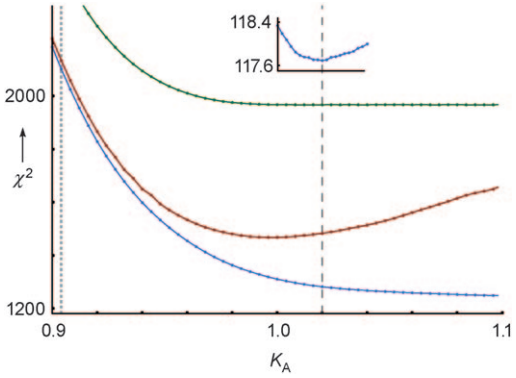

Figure 3. Effects of alignment-tensor scaling on data reproduction with models S (green), 1D GAF/S (red), and 3D GAF (blue). K_A was varied over the range 0.9–1.1 with respect to the minimum determined by 1D GAF/S analysis. The minimum determined from cross-validated 3D GAF analysis (inset) best represents the true alignment tensor.

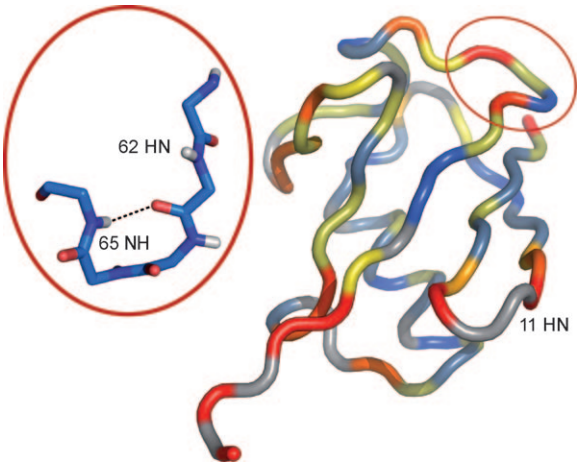

Figure 4. H^N–N order parameters (S^2_{RDC}) from 3D GAF analysis of ubiquitin. Scale from dark blue (1.0) to 0.50 (dark red) via green, yellow, orange. Grey: not determined. Insert: Turn region 62–65 showing higher-amplitude γ/α motions and the hydrogen bond across this turn.

in both experimental and simulated cases (see the inset in Figure 3 and the Supporting Information). The dotted line (on the left-hand side of the graph in Figure 3) shows the alignment tensor determined on the assumption of no dynamics. This alignment tensor is significantly smaller and leads to poorer reproduction of the data.

The robustness of the SF GAF approach was tested by cross-validation: a) Data from each alignment medium were removed for 24 separate analyses, and RDCs were back-calculated by using static and 3D GAF models and appropriate alignment tensors for the specific models; b) two N–H^N RDCs were randomly removed from each peptide plane and back-calculated. In both cases, the χ^2 value was considerably lower for 3D GAF analysis than for the static model (1.1 compared to 4.2 for (a), and 1.0 compared to 3.7 for (b); see the Supporting Information). This demonstrates that the dynamics determined by 3D GAF analysis are required for correct reproduction of the data, providing significant improvement over an analysis using optimal tensors for a static analysis. Average orientations of internuclear bonds determined by this approach are also very similar to those in the high-resolution NMR structure (PDB code: 1d3z;[29] see the Supporting Information).

Enhanced motions apparent in the turn region of the N-terminal β hairpin (8–12) occur in the time range covered by the MD simulation, but are invisible to spin-relaxation experiments due to the overall tumbling of the molecule (around 4 ns). Dynamics occurring in the region of the turn 62–65 appear to reflect larger γ motions for peptide planes 62 and 65 (Figure 4) and are not present in the simulation. A hydrogen bond present in this turn is one of the weakest detected from $^3J_{N,C'}$ scalar coupling.[30] Peptide planes 64 and 65 show the largest-amplitude (30°) α motions in the protein, suggesting mutually dependent dynamics in this turn.

In conclusion, RDCs from ubiquitin measured in multiple alignment media were analyzed in terms of local backbone dynamics by using a structure-free GAF-based approach. This approach yielded significant improvement in data reproduction over a model supposing no dynamics. The method relies only on experimental RDCs; thus, absolute alignment-tensor information and quantitative internal motional modes and amplitudes are obtained from experimental data alone. Comparison with simulation confirmed that the approach can be used to determine internal mobility on an absolute scale. In common with all RDC-based methods, this approach is insensitive to the potential presence of isotropic internal motions of identical magnitude for all dipolar interactions across the whole protein. Nevertheless, there is close correspondence of the average amount, nature, and distribution of motion derived using the SF GAF model compared to the motion present in a 400 ns MD trajectory of ubiquitin. The method also reproduces SCRM-derived order parameters remarkably well despite the fact that no scaling with respect to the relaxation order parameters is applied in the SF GAF approach. More generally, these methods lay the foundation for a quantitative description of the local dynamic behavior of proteins on a wide range of time scales up to the millisecond on the basis of RDCs alone, provided the protein of interest is amenable to study in a sufficient number of different alignment media.[25] This type of analysis provides important information about protein flexibility and will hopefully improve our understanding of the mechanisms governing molecular recognition and function.

Received: January 25, 2009
Published online: May 4, 2009

Keywords: Gaussian axial fluctuation · molecular dynamics · NMR spectroscopy · protein dynamics · residual dipolar coupling

[1] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard, *Nat. Struct. Biol.* **1997**, *4*, 292–297.
[2] N. Tjandra, A. Bax, *Science* **1997**, *278*, 1111–1116.
[3] J. H. Prestegard, H. M. Al-Hashimi, J. R. Tolman, *Q. Rev. Biophys.* **2000**, *33*, 371–424.
[4] J. R. Tolman, *J. Am. Chem. Soc.* **2002**, *124*, 12020–12031.

[5] J. Meiler, J. J. Prompers, W. Peti, C. Griesinger, R. Brüschweiler, *J. Am. Chem. Soc.* **2001**, *123*, 6098 – 6107.
[6] G. Bouvignies, P. Bernado, S. Meier, K. Cho, S. Grzesiek, R. Brüschweiler, M. Blackledge, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13885 – 13890.
[7] G. Bouvignies, P. R. L. Markwick, R. Brüschweiler, M. Blackledge, *J. Am. Chem. Soc.* **2006**, *128*, 15100 – 15101.
[8] L. Yao, B. Vogeli, D. Torchia, A. Bax, *J. Phys. Chem. B* **2008**, *112*, 6045 – 6056.
[9] G. M. Clore, C. D. Schwieters, *J. Am. Chem. Soc.* **2004**, *126*, 2923 – 2938.
[10] O. F. Lange, N.-A. Lakomek, C. Farès, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, B. L. de Groot, *Science* **2008**, *320*, 1471 – 1475.
[11] K. Lindorff-Larsen, R. B. Best, M. A. DePristo, C. M. Dobson, M. Vendruscolo, *Nature* **2005**, *433*, 128 – 132.
[12] S. A. Showalter, R. Brüschweiler, *J. Am. Chem. Soc.* **2007**, *129*, 4158 – 4159.
[13] P. R. L. Markwick, G. Bouvignies, M. Blackledge, *J. Am. Chem. Soc.* **2007**, *129*, 4724 – 4730.
[14] N. A. Lakomek, K. F. A. Walter, C. Farès, O. F. Lange, B. L. de Groot, H. Grubmüller, R. Brüschweiler, A. Munk, S. Becker, J. Meiler, C. Griesinger, *J. Biomol. NMR* **2008**, *41*, 139 – 155.
[15] R. Brüschweiler, P. E. Wright, *J. Am. Chem. Soc.* **1994**, *116*, 8426 – 8427.
[16] T. Bremi, R. Brüschweiler, *J. Am. Chem. Soc.* **1997**, *119*, 6672 – 6673.
[17] P. R. L. Markwick, R. Sprangers, M. Sattler, *Angew. Chem.* **2005**, *117*, 3296 – 3301; *Angew. Chem. Int. Ed.* **2005**, *44*, 3232 – 3237.
[18] K. B. Briggman, J. R. Tolman, *J. Am. Chem. Soc.* **2003**, *125*, 10164 – 10165.
[19] J. R. Tolman, K. Ruan, *Chem. Rev.* **2006**, *106*, 1720 – 1736.
[20] M. Ottiger, A. Bax, *J. Am. Chem. Soc.* **1998**, *120*, 12334 – 12341.
[21] P. Bernadó, M. Blackledge, *J. Am. Chem. Soc.* **2004**, *126*, 4907 – 4920.
[22] J. C. Hus, R. Brüschweiler, *J. Biomol. NMR* **2002**, *24*, 123 – 132.
[23] P. J. Rousseeuw, C. Croux, *J. Am. Stat. Assoc.* **1993**, *88*, 1273 – 1283.
[24] L. Yao, B. Vögeli, J. Ying, A. Bax, *J. Am. Chem. Soc.* **2008**, *130*, 16518 – 16520.
[25] G. Bouvignies, P. R. L. Markwick, M. Blackledge, *Proteins Struct. Funct. Bioinf.* **2008**, *71*, 353 – 363.
[26] H. Motulsky, A. Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*, Oxford University Press, Oxford, **2004**.
[27] S. F. Lienin, T. Bremi, B. Brutscher, R. Brüschweiler, R. R. Ernst, *J. Am. Chem. Soc.* **1998**, *120*, 9870 – 9879.
[28] N. A. Lakomek, T. Carlomagno, S. Becker, C. Griesinger, J. Meiler, *J. Biomol. NMR* **2006**, *34*, 101 – 115.
[29] G. Cornilescu, J. L. Marquardt, M. Ottiger, A. Bax, *J. Am. Chem. Soc.* **1998**, *120*, 6836 – 6837.
[30] F. Cordier, S. Grzesiek, *J. Mol. Biol.* **2002**, *317*, 739 – 752.

