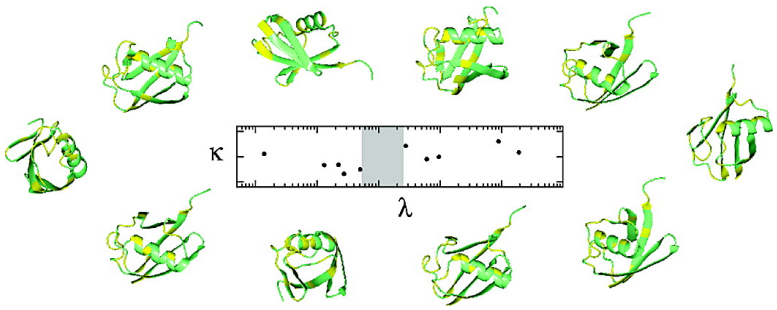
Communication

Self-Consistency Analysis of Dipolar Couplings in Multiple Alignments of Ubiquitin

Jean-Christophe Hus, Wolfgang Peti, Christian Griesinger, and Rafael Brschweiler

J. Am. Chem. Soc., **2003**, 125 (19), 5596-5597• DOI: 10.1021/ja029719s • Publication Date (Web): 19 April 2003



More About This Article

J|A|C|S
COMMUNICATIONS

Published on Web 04/19/2003

Self-Consistency Analysis of Dipolar Couplings in Multiple Alignments of Ubiquitin

Jean-Christophe Hus,[†] Wolfgang Peti,[‡] Christian Griesinger,[§] and Rafael Brüschweiler*,[†]

Carlson School of Chemistry and Biochemistry, Clark University, Worcester, Massachusetts, Department of Molecular Biology, The Scripps Research Institute, La Jolla, California, and Max Planck Institut für Biophysik, Göttingen, Germany

Received December 12, 2002; E-mail: bruschweiler@nmr.clarku.edu

NMR residual dipolar couplings (RDCs) provide useful structural and dynamic information on biomolecules in solution.[1,2] Weak alignment is caused by specific and nonspecific interactions between the solute macromolecule and the alignment medium, which can be described by an alignment tensor.[3] For the interpretation of RDCs it is essential to know whether the solute's structure and internal dynamics are identical in all media (homogeneous behavior) or whether they differ (heterogeneous behavior).[4−6] Heterogeneities can be induced by the alignment media themselves or by variable pH or temperature. Recently, an analysis method, termed SECONDA,[6] has been proposed that allows the assessment of the self-consistency of RDC data collected in multiple alignments.[7,8] It is applied here to RDCs of backbone N−H vectors of the protein ubiquitin measured for 10 different alignment media.[9]

The weighted covariance matrix $\boldsymbol{C}$ of RDCs $D_i^{(k)}$ that belong to N vectors i measured in M alignments k is[6]

$$C_{ij} = \frac{1}{M-1}\sum_{k=1}^{M}(D_i^{(k)} - \langle D_i\rangle)(D_j^{(k)} - \langle D_j\rangle)/\sigma_k^2 \qquad (1)$$

$\langle D_i\rangle$ is the average of couplings $D_i^{(k)}$ over the M alignments and σ_k^2 is the variance of the couplings measured in medium k. A principal component analysis is applied to $\boldsymbol{C}$ by diagonalization, $\boldsymbol{C}|q\rangle = \lambda_q|q\rangle$, where λ_q are eigenvalues to the N normalized eigenmodes $|q\rangle$. The five largest λ_q and their eigenmodes encode the average N−H vector orientations,[12] whereas the appearance of additional nonzero eigenvalues provides direct evidence for inconsistencies, which are structural or dynamic heterogeneities or experimental noise. The gap between the fifth and the sixth largest eigenvalue is given by $\rho = \lambda_5/\lambda_6$, which is a measure of the degree of consistency in the data. The percentage of dipolar vectors that are significantly affected by eigenmode $|q\rangle$ is expressed as collectivity κ_q.[6]

Of a theoretical maximum of 720 N−H RDCs 624 RDCs could be measured (no cross-peak overlaps or line broadening). Because eq 1 is sensitive to missing RDCs, one can (i) apply eq 1 exclusively to N−H groups with measurable RDCs in all media, (ii) exclude certain alignment media for which several RDCs are missing, or (iii) back-calculate missing RDCs from a structural model. All three strategies were applied with the results displayed in Figure 1.

(i) RDCs in all 10 media are available for 32 N−H groups. For these RDCs matrix $\boldsymbol{C}$ of eq 1 was determined and a principal component analysis was applied. The resulting κ,λ distribution is shown in Figure 1A. Nine eigenvalues are nonzero with the five largest eigenvalues separated by a gap $\rho = 11.5$ from the following four eigenvalues that reflect inconsistent behavior. The cumulative effect of the heterogeneous modes on each N−H bond vector j can be expressed as $a_j^2 = \sum_q \lambda_q |q\rangle_j^2$, where the sum includes all heterogeneous modes $|q\rangle$ with elements $|q\rangle_j$. A small a_j^2 value indicates that the RDCs of vector j exhibit homogeneous behavior, while a large a_j^2 value, which is present for Gln 2, Thr 55, and Gln 62, reflect heterogeneity or noise.

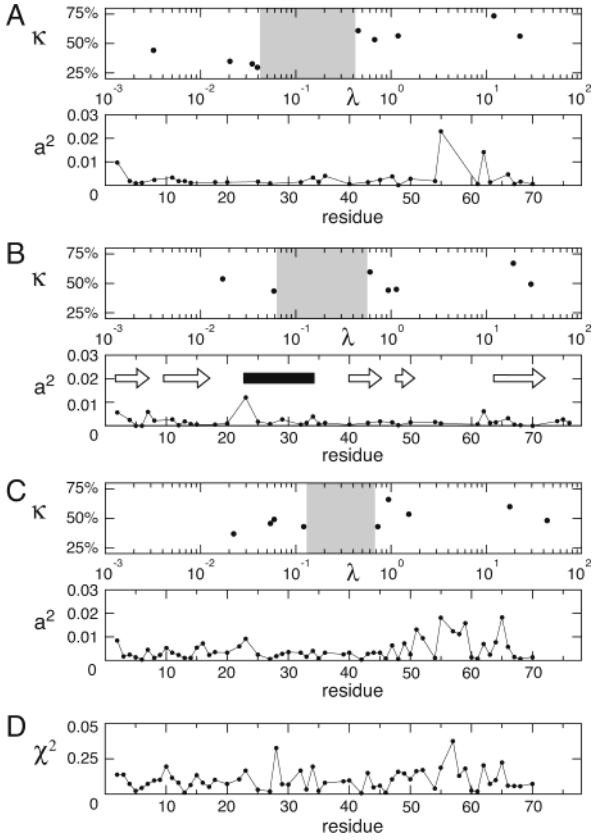


Figure 1. Self-consistency analysis (SECONDA) of ^{15}N−^{1}H dipolar couplings of ubiquitin in up to 10 different alignment media. Panel A: κ,λ plot using 32 N−H vectors with RDC data in all 10 media.[9] The lower panel shows the cumulative sum of heterogeneous modes for individual residues. Panel B: SECONDA analysis of 41 N−H vectors in eight media. Panel C: SECONDA analysis of 58 residues in 10 media, where missing RDCs were back-calculated from average N−H orientations. Panel D: χ_j^2 of eq 2 as a function of residue number.

(ii) Exclusion of the CHAPSO/DLPC/SDS and CHAPSO/DLPC/CTAB (4%) media allows application of SECONDA to 41 N−H vectors that include the highly mobile C-terminal residues 74−76 (Figure 1B). A gap $\rho = 10.2$ is found, which is only slightly smaller than in analysis (i). Gln 2, Thr 7, Ile 23, and Gln 62 exhibit inconsistent behavior. Except for Thr 7, these residues were previously found to show increased mobility in ^{15}N relaxation

[†] Clark University.
[‡] The Scripps Research Institute.
[§] Max Planck Institut für Biophysik.

10.1021/ja029719s CCC: $25.00

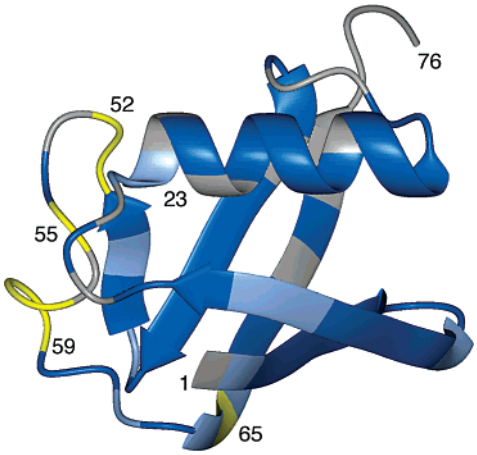


Figure 2. Ubiquitin backbone color-coded according to the level of self-consistency of N−H RDCs (dark blue: high-level of self-consistency; light blue: slightly inconsistent behavior; yellow: significant inconsistencies; gray: not determined).

experiments.[13] The mobile C terminus and the slowly exchanging Asn 25 display homogeneous behavior.

(iii) Missing RDCs were determined from a static structural model, which was obtained by simultaneous determination of the 10 alignment tensors and the N−H vector orientations using a fitting procedure that minimizes $\chi^2 = \sum_j \chi_j^2$ where

$$\chi_j^2 = \sum_{k=1}^{M} (D_{j,\mathrm{calc}}^{(k)} - D_{j,\mathrm{exp}}^{(k)})^2/\sigma_k^2 \qquad (2)$$

is the weighted square difference between experimental and calculated N−H RDCs of vector j in the M alignments.[14] The minimization was performed using 538 experimental RDCs with initial orientations taken from the X-ray structure.[15] From the fitted N−H vector orientations, 42 missing RDCs were then back-calculated and used, together with the experimental RDCs, as input for SECONDA. The resulting κ,λ distribution is shown in Figure 1C exhibiting a gap of 6.0. Because back-calculated RDCs assume a static structure, inconsistencies can also reflect dynamics. In Figure 1D, the best χ_j^2 values are plotted showing a similar trend as the cumulative a_j^2 values in Figure 1C. A general increase of a_j^2 and χ_j^2 is visible between residues 53 and 65, which comprises a large irregularly shaped loop that is (except Gln 62) static in the NMR relaxation experiments.[13] The fitted N−H vector orientations deviate on average by 8.8° from the N−H vectors determined from the X-ray structure[15] and by 6.4° from the NMR structure,[16] which is in agreement with previous results.[7,8] In Figure 2, each residue of ubiquitin is color-coded according to its level of consistency found in SECONDA analyses (i−iii).

A quantitative breakdown of inconsistencies into heterogeneous behavior and noise is not straightforward. The SECONDA analysis, however, allows one to obtain an upper noise limit, assuming that no heterogeneities are present. By adding various amounts of Gaussian noise to RDCs back-calculated from the fitted N−H vector orientations, we find that a gap of 6.0 corresponds to ~10% noise. This amount exceeds noise estimates obtained by repeating RDC measurements.[7,16] Moreover, Gaussian noise leads to a dispersion in λ that is much smaller than what is found in Figure 1.[6] This suggests that heterogeneous or dynamic behavior is at least partially responsible for these inconsistencies.

This conclusion is further supported when considering the secondary structures of inconsistent residues. In the three analyses (i−iii), significant inconsistencies are only found for residues that belong neither to the central α-helix nor to β-strands. Residues that exhibit the highest level of inconsistency are clustered in the loop region 51−63. If structural heterogeneity is the only factor, the χ_j^2 values (Figure 1D) correspond to changes in the N−H orientations of the order of 7°, which slightly exceeds the amount of "structural noise" estimated for RDC-refined NMR structures.[17]

The SECONDA results suggest that the regular secondary structural elements of ubiquitin behave highly homogeneously across the 10 alignment media, while small heterogeneities are likely to be present in the loop region 51−63. SECONDA allows one to sensitively monitor structural-dynamic changes upon pH variation. For instance, heterogeneities in loop 8−12 are observed when RDC data in polyacrylamide gel at a more basic pH 7.5 (Sass and Grzesiek, unpublished data) are included.

Acknowledgment. We thank Hans-Jürgen Sass and Stephan Grzesiek for providing unpublished RDC data of ubiquitin aligned in polyacrylamide gel and Theresa Carlomagno and Dirk Lennartz for helpful discussion. This work was supported by NSF Grant MCB-0211512 (to R.B.).

Supporting Information Available: Three tables with RDCs and alignment tensors and figure showing the noise-dependence of ρ (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127−174.
(2) Prestegard, J. H.; Kishore, A. I. *Curr. Opin. Chem. Biol.* **2001**, *5*, 584−590.
(3) Saupe, A. *Angew. Chem., Int. Ed. Engl.* **1968**, *7*, 97−112.
(4) Goto, N. K.; Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Mol. Biol.* **2001**, *308*, 745−764.
(5) Barbieri, R.; Bertini, I.; Lee, Y.-M.; Luchinat, C.; Velders, A. H. *J. Biomol. NMR* **2002**, *22*, 365−368.
(6) Hus, J.-C.; Brüschweiler, R. *J. Biomol. NMR* **2002**, *24*, 123−132.
(7) Peti, W.; Meiler, J.; Brüschweiler, R.; Griesinger, C. *J. Am. Chem. Soc.* **2002**, *124*, 5822−5833.
(8) Tolman, J. R. *J. Am. Chem. Soc.* **2002**, *124*, 12020−12030.
(9) Backbone $^{15}N-^{1}H$ RDCs of ubiquitin measured in different labs in the following alignment media were used: DHPC/DMPC,[10] DHPC/DMPC/CTAB,[10] purple membranes,[7] CHAPSO/DLPC/SDS (10:50:1; 5%),[7] CHAPSO/DLPC (1:5; 5%),[7] CHAPSO/DLPC/CTAB(10:50:1; 4%),[7] CHAPSO/DLPC/CTAB (10:50:1; 5%),[7] Helfrich phase,[7] *n*-dodecyl-penta-(ethylene glycol)/*n*-hexanol,[7] and Pf-1 phages.[11] All data were measured at 303 K and pH 6.5, except for the RDCs in DHPC/DMPC and DHPC/DMPC/CTAB media that were measured at 308 K.
(10) Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 12334−12431.
(11) Hus, J.-C.; Prompers, J. J.; Brüschweiler, R. *J. Magn. Reson.* **2002**, *157*, 119−123.
(12) Hus, J.-C.; Brüschweiler, R. *J. Chem. Phys.* **2002**, *117*, 1166−1172.
(13) (a) Schneider, D. M.; Dellwo, M. J.; Wand, A. J. *Biochemistry* **1992**, *31*, 3645−3652. (b) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax A. *J. Am. Chem. Soc.* **1995**, *117*, 12562−12566. (c) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870−9879.
(14) Residues 21, 24, 26, 31, 41, 53, 56, 69, 71, and 72 were discarded because of spectral overlaps and line broadening effects, while the mobile C-terminal residues 73−76 were excluded because they cannot be adequately represented by a static structure.
(15) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544. PDB file 1ubq.
(16) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837. PDB file 1d3z.
(17) Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2002**, *23*, 127−137.

JA029719S