J|A|C|S
ARTICLES
Published on Web 02/12/2004

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?

G. Marius Clore*,† and Charles D. Schwieters*,‡

Contribution from the Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, and Division of Computational Bioscience, Building 12A, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892-5624

Received September 23, 2003; E-mail: mariusc@intra.niddk.nih.gov; charles.schwieters@nih.gov

Abstract: The magnitude of backbone internal motions in the small protein ubiquitin that needs to be invoked to account for dipolar coupling data measured in multiple alignment media is investigated using an intuitively straightforward approach. This involves simultaneous refinement of the coordinates (against NOE, torsion angle, and dipolar coupling restraints) and optimization of the magnitudes and orientations of the alignment tensors by means of torsion angle simulated annealing and Cartesian space minimization. We show that N−H dipolar couplings in 11 different alignment media and N−C′, H_N−C′, and Cα−C′ dipolar coupling in two alignment media can be accounted for, at approximately the level of uncertainty in the experimental data, by a single structure representation. Extension to a two-member ensemble representation which provides the simplest description of anisotropic motions in the form of a two-site jump model (in which the overall calculated dipolar couplings are the averages of the calculated dipolar couplings of the individual ensemble members), results in modest, but significant, improvements in dipolar coupling *R*-factors for both the working set of couplings used in the refinement and for the free cross-validated set of Cα−Hα dipolar couplings recorded in two alignment media. Extensions to larger ensemble sizes do not result in any *R*-factor improvement for the cross-validated Cα−Hα dipolar couplings. With a few notable exceptions, the amplitudes of the anisotropic motions are small, with S^2(jump) order parameters ≥0.8. Moreover, the structural impact of those few residues that do exhibit larger amplitude motions (S^2(jump) ranging from 0.3 to 0.8) is minimal and can readily be accommodated by very small backbone atomic rms shifts (<0.5 Å) because of compensatory changes in ϕ and ψ backbone torsion angles. In addition, evidence for correlated motions of N−H bond vectors is observed. For most practical applications, however, refinement of NMR structures against dipolar couplings using a single structure representation is adequate and will not adversely impact coordinate accuracy within the limits of the NMR method.

Introduction

Protein dynamics is crucial to protein function and has been extensively studied using a variety of experimental approaches, including diffraction methods, solution state spectroscopy (NMR, fluorescence, and optical spectroscopy), and solid-state NMR spectroscopy, as well as by theoretical and computational methods such as molecular dynamics simulations (see ref 1 for some reviews). The general picture of globular proteins that emerges from NMR relaxation studies, X-ray crystallography, and molecular dynamics simulations is one in which backbone motions on a time scale <10 ns, with the obvious exception of flexible loop regions, are generally of small overall amplitude, corresponding to values of ∼0.85 for the generalized order parameter S^2 derived from NMR relaxation studies. Larger motions can occur on slower time scales, but generally represent rare events (such as local unfolding and solvent access for backbone amide exchange and aromatic ring flipping).

Partial alignment of macromolecules in a magnetic field using dilute liquid crystalline media[2] has offered a simple means

† National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

‡ Center for Information Technology, National Institutes of Health.

(1) (a) Wagner, G. *Q. Rev. Biophys.* **1983**, *16*, 1−57. (b) Torchia, D. A. *Annu. Rev. Biophys. Bioeng.* **1984**, *13*, 125−144. (c) Ringe, D.; Petsko, G. A. *Prog. Biophys. Mol. Biol.* **1985**, *45*, 197−235. (d) Petsko, G. A.; Karplus, M. *Nature* **1990**, *347*, 631−639. (e) Englander, S. W.; Sosnick, T. R.; Englander, J. J.; Mayne, L. *Curr. Opin. Struct. Biol.* **1996**, *6*, 18−23. (f) Kay, L. E. *Nat. Struct. Biol.* **1998**, *5*, S513−S517. (g) Ishima, R.; Torchia, D. A. *Nat. Struct. Biol.* **2000**, *7*, 740−743. (h) Wand, A. J. *Nat. Struct. Biol.* **2001**, *8*, 926−931. (i) Palmer, A. G., III. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 129−155l. (j) Karplus, M.; McCammon. J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652. (k) Frauenfleder, H.; McMahon, B. H.; Fenimore, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8615−8617.

(2) (a) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114. (b) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571−10572. (c) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065−1074. (d) Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371−424. (e) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127−174.

of measuring residual dipolar couplings which afford unique long-range orientational information for structure determination.[3] Local motion on a time scale up to ~1 ms results in partial averaging and reduction of the observed residual dipolar couplings which, in the case of small amplitude isotropic motions, scale with the order parameter $\boldsymbol{S}$ (as opposed to $\boldsymbol{S}^2$ in the case of relaxation measurements).[2a] A series of recent studies using residual dipolar coupling measurements has proposed the existence of ubiquitous large amplitude collective motions that are not simply present transiently but are present at all times on a time scale ranging from nanoseconds to ~1 ms.[4−8] The first such study, based on magnetically induced partial alignment of paramagnetic myoglobin, suggested the presence of deviations in axis orientations of several helices by ≥25° from their averaged positions to account for the observation that the measured N−H_N dipolar couplings were ~27% smaller than those predicted from the static crystal structure and the magnitude of the magnetic susceptibility tensor determined from paramagnetic shifts.[4] However, it was also noted that if the magnetic susceptibility tensor is derived from the dipolar couplings themselves, excellent agreement between all observed and calculated dipolar couplings is obtained on the basis of the static crystal structure without the need to invoke large amplitude motions.[9] More recent work[5−8] has focused extensively on the small 76-residue protein ubiquitin which has long served as a model system for the development of new NMR methodologies and for the NMR analysis of fundamental properties of proteins.

In an extensive study based on N−H dipolar couplings measured in 11 different liquid crystalline media, it was suggested that the single α-helix of ubiquitin undergoes a concerted rigid body anisotropic motion with amplitudes of ~±21° and ~±12° along two axes mutually orthogonal to the long axis of the α-helix, with the first parallel and the second orthogonal to the underlying β-sheet.[7] However, the α-helix lies on top of a four-stranded sheet, and its position is governed by interdigitation of numerous side chains involving residues of the helix and the underlying sheet.[10] The latter interactions constitute the hydrophobic core of ubiquitin. Thus, any concerted motion of the helix would require corresponding concerted motions of a large number of underlying side chains. It is interesting to note, in this regard, that in another study on the small B3 domain of Streptococcal protein G, which has the same protein fold as ubiquitin, the backbone dipolar couplings (N−H_N, N−C′, and Cα−C′) measured in five different media could be reasonably well accounted for by a single static structure with minimal deviations in idealized covalent geometry.[11] If such large scale concerted motions do indeed occur in ubiquitin, it would suggest that crystallization selects only a single conformation which fortuitously corresponds very closely to the average conformation observed in solution[12] (since backbone rms difference between the X-ray and NMR structures is only ~0.4 Å).

A number of different approaches have been suggested for deriving backbone motional information from dipolar couplings. The first makes use of a single alignment medium and a set of seven heteronuclear backbone dipolar couplings to analyze each peptide fragment individually.[5a] While elegant in principle, the method in practice is highly sensitive to coordinate errors since the number of dipolar couplings measured for each peptide fragment analyzed is still small.[13] The second approach consists of a "model free" analysis of dipolar couplings measured in multiple alignment media and involves the extraction of spherical harmonics and effective vector orientations.[6,7] This approach appears to be susceptible to cumulative errors and also depends on the accuracy of the coordinates used which can substantially impact the values of the alignment tensor[13] critical to the analysis. In all the various analyses, no attempt has been made to use some independent measurement to verify the requirement of internal motion by cross-validation.

In the present article, we set out to determine whether the N−H dipolar couplings measured in the 11 different alignment media employed by Peti et al.[7a] require one to invoke the presence of backbone internal motion, and if so, what is the minimum magnitude of the motion needed to account for the experimental data. The approach we employ makes use of a straightforward, intuitively simple approach based on the application of Occam's razor. Specifically, we refine the solution structure of ubiquitin using nuclear Overhauser enhancement (NOE) data,[12] N−H dipolar couplings measured in 11 different alignment media,[7] and N−C′, H_N−C′, and Cα−C′ dipolar couplings measured in two different alignment media[12] by simultaneously refining the coordinates of the protein and the orientation and magnitude of the alignment tensors, using a combination of torsion angle simulated annealing and Cartesian space minimization.[14] The results are then cross-validated[15] against a number of independent measurements which include Cα−Hα dipolar couplings in two alignment media,[12] H_N−Hα dipolar couplings in one alignment medium,[2a,16] and six different sets of three-bond ^3J scalar backbone couplings.[17] We first show that very significant improvements in agreement between observed and calculated dipolar couplings, approximately to the level of experimental uncertainty, are readily obtained upon refinement of a single model representation. Further modest improvements in agreement, as well as improvements in the cross-validated Cα−Hα dipolar coupling terms, are obtained

(3) (a) Tjandra, N.; Omichinski, J. G.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *Nat. Struct. Biol.* **1998**, *4*, 732−739. (b) Bewley, C. A.; Gustfason, K. R.; Boyd, M. R.; Covell, D. G.; Bax, A.; Clore, G. M.; Gronenborn, A. M. *Nat. Struct. Biol.* **1998**, *5*, 571−578. (c) Clore, G. M.; Starich, M. R.; Bewley, C. A.; Cai, M.; Kuszewski, J. *J. Am. Chem. Soc.* **1999**, *121*, 6513−6514. (d) Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *97*, 9021−9025. (e) Skrynnikov, N. R.; Goto, N. K.; Yang, D.; Choi, W. Y.; Tolman, J. R.; Mueller, G. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *295*, 1265−1273. (f) Goto, N. K.; Skrynnikov, N. R.; Dalhquist, F. W.; Kay, L. E. *J. Mol. Biol.* **2001**, *308*, 745−764. (g) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2003**, *125*, 2902−2912.
(4) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Nat. Struct. Biol.* **1997**, *4*, 292−297.
(5) (a) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 1416−1424. (b) Tolman, J. R. *J. Am. Chem. Soc.* **2002**, *124*, 12020−12030. (c) Briggman, K. B.; Tolman, J. R. *J. Am. Chem. Soc.* **2003**, *125*, 10164−10165.
(6) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098−6107.
(7) (a) Peti, W.; Meiler, J.; Brüschweiler, R.; Griesinger, C. *J. Am. Chem. Soc.* **2002**, *124*, 5822−5833. (b) Meiler, J.; Peti, W.; Griesinger, C. *J. Am. Chem. Soc.* **2003**, *125*, 8072−8073.
(8) Hus, J.-C.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2003**, *125*, 5596−5597.
(9) Bax, A.; Tjandra, N. *Nat. Struct. Biol.* **1997**, *4*, 254−256.
(10) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544.
(11) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179−9191.
(12) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837.
(13) Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2002**, *23*, 127−137.
(14) Schwieters, C. D.; Clore, G. M. *J. Magn. Reson.* **2001**, *152*, 288−302.
(15) Clore, G. M.; Garrett, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 9008−9012.
(16) Tjandra, N.; Marquardt, J. L.; Clore, G. M. *J. Magn. Reson.* **2000**, *142*, 393−396.
(17) Hu, J.-S.; Bax, A. *J. Am. Chem. Soc.* **1997**, *119*, 6360−6368.

using a multiple structure representation with an ensemble size of two. In this representation, which provides the simplest representation of anisotropic motion, the overall calculated dipolar couplings for the ensemble are given by the average of the calculated dipolar couplings of the individual members of the ensemble. While some additional improvement in the fitted dipolar couplings can be obtained by increasing the ensemble size further, no improvement is obtained in the cross-validated terms. The analysis suggests that the order parameter, S^2(jump), for a two-site jump is generally greater than 0.8 with a few notable exceptions where larger degrees of motion seem to occur. These larger motions, however, are highly local and can be achieved with minimal perturbations (~0.5 Å) in the overall backbone coordinates.

Methods

General Computational Tools. All dynamics and minimization calculations were carried out using the IVM Internal Variable Module[14] within the molecular structure determination package Xplor-NIH,[18] incorporating the various new features related to ensemble averaging and dipolar couplings described below. Structures were displayed using the program VMD-XPLOR.[19] (Note that all the figures display residues 2−72, since the C-terminal four residues are disordered in solution.)

Ensemble Averaging. The simplest approach to take into account both molecular motions and structural heterogeneity in NMR structure refinement is to model the system using an ensemble of molecular structures. From a purely heuristic perspective, an ensemble size of $N_e = 2$ represents the simplest description of anisotropic motion (see further discussion of this representation in the context of a description of local motions in the Results section).

To carry out structure refinement using ensemble averaging we have introduced an efficient and easy-to-use ensemble feature into Xplor-NIH.[18] Ensemble size is controlled by changing a single number in the input file: alternate structure files for each size are not necessary, and no special care needs to be taken such that the ensemble members do not interact. In addition, the calculation time scales approximately as the inverse of the number of CPUs, up to N_e, on multiple processor, shared memory computer hardware.

Ensemble-averaged quantities are denoted inside angle brackets and are calculated as:

$$\langle x \rangle_e = \sum_i \Gamma_i x_i \tag{1}$$

where x_i is the value of quantity x in ensemble member i and Γ_i is the weight on the ith member. In the current study, Γ_i is taken as $1/N_e$, where N_e is the ensemble size.

General Potential Terms. A number of Xplor-NIH[18] potential terms are used to maintain proper covalent geometry and to prevent atomic overlap. The contribution of a particular Xplor energy term E_α to the overall energy is just its ensemble averaged value: $\langle E_\alpha \rangle_e$. Since the NMR experiments measure ensemble-averaged quantities, simple averaging is not appropriate, and for these terms ensemble averaging must be included in the calculated observables as shown below.

The Residual Dipolar Coupling Term. The dipolar coupling $\delta_{m,n}$ between a pair of atoms described by indices m and n is given by:[1−3]

$$\delta_{m,n} = D_a(3u^2_{m,n,z} - 1) + \frac{3}{2}D_a\eta(u^2_{m,n,x} - u^2_{m,n,y}) \tag{2}$$

where $u_{m,n,x}$, $u_{m,n,y}$, and $u_{m,n,z}$ are the projections of the unit vector in the direction $q_n - q_m$, along the x, y, and z axes of the alignment tensor orientation, respectively. In the definitions given here and below, q_i denotes the Cartesian position of atom i. D_a and D_r are, respectively, the axial and rhombic tensor components, with the rhombicity η defined as D_r/D_a. The tensor is represented using six pseudo atoms. Four atoms (X, Y, Z, and O) represent the orientation of the tensor principal axes (with atom O at the origin), and two atoms, P_1 and P_2, are used to represent D_a and η, respectively. The representation of the tensor magnitude is:

$$D_a = D_{amax} \cos \theta_1 \tag{3}$$

where D_{amax} is the maximum allowed value of D_a and θ_1 is the axial angle between the projection of the $q_{P_1}-q_O$ vector on the $X-O-Y$ plane and the q_X-q_O vector and takes values between 0 and 2π. The rhombicity is written as:

$$\eta = \frac{2}{3}\sin \phi_2 \tag{4}$$

where ϕ_2 is the azimuthal angle between the vectors q_Z-q_O and $q_{P_2}-q_O$ and takes values between 0 and π.

The associated restraint energy term, E_{RDC}, is written as:

$$E_{RDC} = w_{RDC}V_{pQuad}(\delta_{mn} - \delta^{obs}_{mn}; 0, 0) \tag{5}$$

where w_{RDC} is a constant weighting factor for this potential term and δ^{obs}_{mn} is the observed dipolar coupling value. The piecewise quadratic potential function is defined as:

$$V_{pQuad}(x; x^+, x^-) = \begin{cases} (x - x^+)^2 & \text{if } x > +x^+ \\ (x + x^-)^2 & \text{if } x < -x^- \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Note that in the context of the present work x^+ and x^- are zero for E_{RDC}, and thus, the potential is a simple harmonic oscillator. The more general form of the equation is used below in other energy terms.

Each ensemble member is allowed to take its own value of D_a and η. This allows for alignment tensors consistent with ensemble members having different shapes.[20] In this work, a single tensor orientation is used for all ensemble members. The orientational differences are then represented by overall rotations of one ensemble member relative to another. One expects the D_a and η of ensemble members to differ by only a small amount, and therefore, we introduce energy terms to restrain the spread of D_a and η within $\pm\Delta D_a$ and $\pm\Delta\eta$, respectively:

$$E_{\Delta D_a} = w_{D_a}\langle V_{pQuad}(D_a - \langle D_a \rangle_e; \Delta D_a, \Delta D_a)\rangle_e \tag{7}$$

and

$$E_{\Delta\eta} = w_\eta\langle V_{pQuad}(\eta - \langle \eta \rangle_e; \Delta\eta, \Delta\eta)\rangle_e \tag{8}$$

where w_{Da} and w_η are weighting factors, and ΔD_a and $\Delta\eta$ denote the allowed deviation from the ensemble averaged quantities. When ΔD_a and $\Delta\eta$ are set to zero, all members of the ensemble are restrained to have the same D_a and η values.

This study includes dipolar coupling measurements involving pairs of the following atom types: NH, CC, CH, and HH. For a given orienting medium, a single tensor should be used for all experiments. Equation 2 is used for the NH dipolar coupling experiments, and non-NH dipolar couplings are normalized by multiplying eq 2 by the prefactor

$$\frac{\gamma_m\gamma_n r^3_{NH}}{\gamma_N\gamma_H r^3_{mn}} \tag{9}$$

(18) Schwieters, C. D.; Kuszewski, J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 66−74.
(19) Schwieters, C. D.; Clore, G. M. *J. Magn. Reson.* **2001**, *139*, 239−244.
(20) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791−3792.

where γ_m is the gyromagnetic ratio of atom m and r_{mn} is the internuclear distance between atoms m and n. Since the HH distance does not correspond to a vector of fixed length and can thus vary, the $1/r_{HH}^3$ HH dependence must be explicitly included in the dipolar coupling calculation.[16]

The NOE-Derived Interproton Distance Restraint Term. For each NOE assignment, we used the following two NOE energy terms:

$$E_{NOE}^{(ens)} = w_{NOE}^{(ens)} V_{pQuad}(r_{NOE}^{(ens)} - r_{NOE}^{(obs)}; r_{NOE}^{+}, r_{NOE}^{-}) \quad (10)$$

and

$$E_{NOE}^{(struct)} = w_{NOE}^{(struct)} \langle V_{pQuad}(r_{NOE}^{(struct)} - r_{NOE}^{(obs)}; r_{NOE}^{+}, r_{NOE}^{-}) \rangle_e \quad (11)$$

where w_{NOE} is a constant weighting factor, $r_{NOE}^{(obs)}$ is the observed NOE distance restraint, $r_{NOE}^{\pm}$ denotes uncertainties in this distance, and $r_{NOE}^{(ens)}$ and $r_{NOE}^{(struct)}$ are NOE distances calculated by different averaging methods. The definitions of these distances are:

$$r_{NOE}^{(ens)} = [\langle \sum_{ij} |q_i - q_j|^{-6} \rangle_e]^{-1/6} \quad (12)$$

and

$$r_{NOE}^{(struct)} = [\sum_{ij} |q_i - q_j|^{-6}]^{-1/6} \quad (13)$$

where the ij sum is over all atom pairs associated with the given NOE restraint. (The ij sum, for example, applies to methyl groups and nonstereoassigned methylene protons.) Note carefully in eq 12 that an ensemble average is performed before the fractional power is applied, and that this quantity corresponds to the observed NOE distance of an ensemble of structures. Equation 13 contains no such average and was used in the annealing protocol to provide an alternate set of forces.

The Relative Atomic Position (RAP) Term. The RAP term restrains the atomic positions in each ensemble member such that they do not stray too far from their respective ensemble-averaged positions:

$$E_{RAP} = w_{RAP} \sum_i \langle V_{pQuad}(|q_i - \langle q_i \rangle_e|; \Delta l_{RAP}, \Delta l_{RAP}) \rangle_e \quad (14)$$

where w_{RAP} is a constant weighting factor, Δl_{RAP} is the allowed distance deviation, and the sum is over all atoms to be restrained. As shown in the Results and Discussion section, when applied to Cα atoms, this term limits the backbone atomic rms differences between members of an ensemble, but only minimally impacts differences in N−H bond vector orientations between members of an ensemble.

The Molecular Shape Term. This term was introduced to prevent excessive rotation and deformation of one ensemble member relative to another. Molecular shape is approximately represented by a massless inertia tensor[21] analogous to the dipolar coupling alignment tensor.

The massless inertia tensor can be written as:

$$T_{shape} = \sum_i \begin{pmatrix} y_i^2 + z_i^2 & -x_i y_i & -x_i z_i \\ -x_i y_i & x_i^2 + z_i^2 & -y_i z_i \\ -x_i z_i & -y_i z_i & x_i^2 + y_i^2 \end{pmatrix} \quad (15)$$

where x_i, y_i, and z_i are the components of the Cartesian coordinate of atom i. The sum is over all atoms used to define molecular shape.

The associated energy is then defined in terms of the principal values of the shape tensors of ensemble member structures and the ensemble averaged structure:

$$E_{shape} = w_{orient} \langle V_{pQuad}(\alpha^{mag}; \Delta\alpha^{mag}, \Delta\alpha^{mag}) \rangle_e + w_{size} \sum_{k=1}^{3} \langle V_{pQuad}(\lambda_k - \lambda_k^{mean}; \Delta\lambda, \Delta\lambda) \rangle_e \quad (16)$$

where α^{mag} is the magnitude of the rotation of the principal axes of an ensemble member from that of the ensemble average structure, and $(\lambda_k - \lambda_k^{mean})$ is the difference of eigenvalue (principal value) k from the associated value of the ensemble average structure. w_{orient} and w_{size} are weighting factors, and $\Delta\alpha^{mag}$ and $\Delta\lambda$ denote the allowed deviation of the values from the mean values.

It should be noted that this shape tensor was found to be insufficient to completely describe protein shape and orientation in this study. This approach describes the protein shape as an ellipsoid. For ubiquitin, this approximation was generally found to be too crude as structures with clearly different shapes (including apparent orientation) were assigned nearly identical shape tensors. Thus, the RAP term (eq 14) was necessary to study ensembles with very similar members.

^3J Coupling Constant Restraint Term. The ^3J coupling constant is defined by the empirical Karplus relationship:[18,22]

$$^3J = A \cos^2(\theta + \theta^*) + B \cos(\theta + \theta^*) + C \quad (17)$$

where A, B, and C are fitted parameters and θ^* is a fixed geometric angular term relating the restrained torsion angle θ to the measured one (e.g., for $^3J_{H_NH\alpha}$, θ^* is +60°, which is a constant relating the backbone torsion angle C′(i − 1)−N(i)−Cα(i)−C′(i) to the torsion angle $H_N(i)$−N(i)−Cα(i)−CαH(i) directly probed by the coupling constant). The energy term associated with this coupling constant is then:

$$E_J = w_J V_{pQuad}(\langle ^3J \rangle_e - ^3J^{(obs)}; 0, 0) \quad (18)$$

where w_J is a weighting factor and $^3J^{(obs)}$ is the observed coupling constant. Although this energy term was not included in the calculations, the rms differences between observed and calculated couplings in the case of calculations using an ensemble size of $N_e \geq 2$ are reported as ensemble averages.

Correlation between Ensemble Members. Once structures are calculated, there arises the question of how similar the ensembles are to one another. Restated, the question is: given two ensembles, how closely do the N−H_N vectors of constituent members line up with those from another ensemble and over what range of residues?

We use an analysis which is appropriate for two-member ensembles in which the structures maintain the same overall fold, but which contain local differences. For each ensemble, mean coordinates are calculated and fit to each other. The rigid-body rotation and translation determined in this fitting procedure are then applied separately to the ensemble members, such that their relative orientation (and the value of the calculated dipolar couplings) is maintained.

Now let v_{ikm} be the unit vector along the N−H_N bond vector of residue k of member m (0 or 1) in the ensemble labeled by i. We assume that each vector can be binned into one of two orientations. For two residues labeled k and l and two ensembles denoted i and j, the ensembles are said to coincide for these residues if $f(i,j;k,l) = 1$, where

$$f(i,j;k,l) = \begin{cases} 1 & \text{if } v_{ik0} \cdot v_{jk0} > v_{ik0} \cdot v_{jk1} \text{ and } v_{il0} \cdot v_{jl0} > v_{il0} \cdot v_{jl1} \\ 1 & \text{if } v_{ik0} \cdot v_{jk0} < v_{ik0} \cdot v_{jk1} \text{ and } v_{il0} \cdot v_{jl0} > v_{il0} \cdot v_{jl1} \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

Thus, for two residues, ensembles coincide if the bond vectors of each member of one ensemble are approximately aligned with those of one member in the other ensemble, so that the ensemble members occupy the same bins in both ensembles.

Averaging over all ensembles, the correlation, C_{kl}, between residues k and l is defined as

$$C_{kl} = -1 + \frac{2}{N_{struct}^2} \sum_{ij} f(i,j;k,l) \quad (20)$$

where i and j are summed over all calculated ensembles and N_{struct} is the number of these ensembles. C_{kl} takes values between 0 and 1, with

0 indicating no correlation. $C_{kl} = 1$ indicates perfect correlation, meaning that the two vectors coincide in all ensembles. By definition $C_{kk} = 1$, and we expect large C_{kl} for small $|k - l|$, as well as for otherwise spatially proximal portions of the structure. Note that using this definition, C_{kl} will tend to zero when it involves bond vectors which take very similar orientations because the bins used in the definition of *f(i,j; k,l)* become indistinguishable, and the average of *f(i,j; k,l)* will tend to 1/2.

Structure Refinement. Refinement made use of a combination of torsion angle dynamics and minimization, followed by Cartesian coordinate minimization. The latter is important since it ensures that the stereochemistry of the backbone is not held rigidly fixed (as it is in torsion angle space), and therefore small deviations from idealized covalent geometry (e.g., in peptide bond planarity) are allowed to occur. The atomic masses are all set equal to 100 amu. The force constants for bonds and angular (angles and improper torsions) terms are set to 1000 kcal mol^{-1} Å^2 and 500 kcal mol^{-1} rad^{-2}, respectively, with the exception of the improper torsions related to the peptide bond, which are set to 250 kcal mol^{-1} rad^{-2}. With these force constants, the maximum observed deviation from peptide bond planarity is 5°, and the average deviation from peptide bond planarity is ≤2.5°. Such deviations are entirely within the range observed in high resolution (≤1.75 Å) crystal structures.[24] The target function used in refinement included the following experimental terms: dipolar coupling,[7,12] NOE-derived interproton distance (final force constant = 30 kcal mol^{-1} Å^{-2}), and χ_1 side-chain torsion angle (final force constant = 200 kcal mol^{-1} rad^{-2}) restraints.[25,26] The final force constant used for the $^1D_{NH}$ dipolar couplings is 1 kcal mol^{-1} Hz^{-2} (with the force constants for the $^1D_{NC'}$, $^1D_{HNC'}$, and $^1D_{CaC'}$ dipolar couplings scaled to that for the $^1D_{NH}$ couplings by factors of 25, 5, and 15, respectively). The nonbonded interactions are described by a quartic van der Waals repulsion term (final force constant of 4 kcal mol^{-1} Å^{-4} with a van der Waals radius scale factor of 0.8),[27] supplemented by a torsion angle database potential of mean force (final weighting factor of 1),[28] an empirical hydrogen-bonding term incorporating both distance and angular dependencies for 36 backbone hydrogen bonds (weighting factor of 500)[29] and a radius of gyration restraint (target value of 11.29 Å given by 2.2 $N^{0.38}$, taking into account residues 1−74 only, and a force constant of 100 kcal mol^{-1} Å^{-2}).[30] For calculations with an ensemble size $N_e \geq 2$, four additional terms are included: the spread terms for D_a (eq 7) and rhombicity (eq 8) with weighting factors of 1×10^5 kcal mol^{-1} Hz^{-2} and 1×10^5, respectively; the RAP term (eq 14) applied to the Cα atoms of residues 2−72 only (force constant of 100 kcal mol^{-1} Å^{-2}); and the shape term (eq 16) applied to atoms outside a sphere of 12 Å from the center of the molecule defined by the Cα atoms of residues 1−72 (force constants of 1 kcal mol^{-1} Å^{-2} for the size component and 1000 kcal mol^{-1} rad^{-2} for the orientation component).

The starting coordinates for refinement are those of the previously determined NMR structure 1D3Z.[12] The atoms describing the dipolar coupling alignment tensor are treated throughout the calculations as follows: atoms *X*, *Y*, *Z*, and *O* representing the tensor orientation are held rigid with respect to each other and allowed only rotational degrees of freedom centered about atom *O*; the P_1 and P_2 atoms representing D_a and η, respectively, were bonded to the O atom and were allowed the appropriate bending motion in addition to the tumbling of the axis atoms. The schedule for the refinement protocol is relatively simple and involves the following steps: 0.2 ps of simulated annealing in torsion angle space with cooling from 400 to 300 K and with a number of force constants (dipolar, NOE, torsion angle, van der Waals repulsion, and torsion angle database) increased geometrically to their final values; Powell minimization in torsion angle space; Powell minimization in Cartesian coordinates. This entire process is repeated 16 times. This protocol provides effective and gentle low-temperature refinement. The calculations are repeated multiple times using different random number seeds for the assignment of initial velocities, resulting in a total of 100 ensembles for each set of calculations.

In the case of calculations with $N_e = 2$, three main sets of conditions were investigated: the most restricted case (2r), a partially restricted case (2pr), and an unrestricted (2u) case. In the case of the 2r set of calculations, the values for the total widths of the flat portion of the piecewise quadratic potential (general formula given by eq 6) are $2\Delta l_{RAP} = 0.5$ Å for the RAP term (eq 14), $2\Delta D_a/D_a = 0$ for D_a (eq 7), and $2\Delta\eta = 0$ for the rhombicity (eq 8). (Note that a width of zero indicates that the potential is represented by a simple harmonic oscillator.) The corresponding values of $2\Delta D_a/D_a$ and $2\Delta\eta$ for the 2pr and 2u calculations are 0.1 and 0.15, respectively; for the 2pr calculations $2\Delta l_{RAP} = 0.5$ Å; for the 2u calculations the RAP term is turned off. A number of additional calculations were also carried out. These include the following combinations of values for $2\Delta l_{RAP}$, $2\Delta D_a/D_a$, and $2\Delta\eta$: 0.5, 0.05, and 0.075 Å; 0.5, 0.1, and 0.15 Å; 1.0, 0, and 0 Å; 1.0, 0.05, and 0.075 Å. The results of these calculations were very similar to those for the 2r and 2pr calculations.

Results and Discussion

Alignment Tensor Orientations in the 11 Media. Figure 1 displays the orientations of the alignment tensors for the 11 different media listed in footnote a to Table 1, relative to the NMR coordinates of ubiquitin (PDB accession code 1D3Z; ref 12). The alignment tensors fall into two main groups: media 1−5 (Figure 1b) and media 6−11 (Figure 1c). The orientation of the principal components of the alignment tensors from the first group is approximately orthogonal (102 ± 8°) to those of

(21) Berardi, R.; Fava, C.; Zannoni, C. *Chem. Phys. Lett.* **1995**, *236*, 462−468.
(22) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870−2871.
(23) Sass, J.; Cordier, F.; Hoffman, A.; Rogowski, M.; Cousin, A.; Omichinski, J. G.; Löwen, H.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 2047−2055.
(24) Karplus, P. A. *Protein Sci.* **1996**, *5*, 1406−1420.
(25) The NOE-derived interproton distance restraints deposited with the PDB coordinate 1D3Z consist of 2872 entries.[12] Many of the restraints are represented by multiple entries depending on the number of times a particular restraint was observed in the various spectra. For example, many entries are present in duplicate corresponding to the two symmetrically related cross-peaks observed in 3D ^{13}C-separated NOE spectra. When the multiple entries are eliminated, the total number of restraints is reduced to 1537.[16] In addition, many of the restraints involve NOEs to stereospecifically assigned β-methylene protons. In keeping with our current practice,[30] we chose to represent pairs of restraints to both methylene protons by a single $\Sigma(r^{-6})^{-1/6}$ sum restraint. This has the advantage of removing potential systematic bias in the form of underestimation in the upper bounds of distance restraints involving one of the protons of a methylene pair as a consequence of spin-diffusion effects. Information content, however, is not lost in the process since the corresponding χ_1 side-chain torsion angle restraints are still present and the torsion angle database potential of mean force ensures that the side chains adopt good stereochemistry within allowed rotamers. The distance restraints in the 1D3Z entry are also classified into rather narrow ranges.[12] We therefore chose to reclassify these into the more conventional four distance range categories corresponding to strong (<2.9 Å), medium (<3.5 Å), weak (<5 Å), and very weak (<6 Å) NOEs. As in the case of the modifications involving β-methylene proton restraints described above, this has the advantage of avoiding some systematic underestimates of upper distance bounds. In our experience, this more conservative treatment of the NOE-derived interproton distance restraints increases the accuracy of the calculated coordinates.[30] The final restraints lists consisted of a total of 1119 distances: 227 intraresidue, 283 sequential ($|i - j| = 1$), 229 short ($1 < |i - j| \leq 5$), and 380 long ($|i - j| > 5$) range distance restraints. Calculations were also performed with the original set of 2872 interproton distance restraints, and while there are very small quantitative differences, the conclusions remain identical: the agreement with both the working and free set of dipolar couplings is essentially the same; similarly, for the calculations with an ensemble size of $N_e = 2$, the calculated S^2(jump) order parameters are essentially unaffected by the NOE restraints list employed.
(26) The χ_1 side-chain torsion angle restraints are those deposited with the 1D3Z NMR coordinates and were derived from heteronuclear ^3J coupling measurements.[12]
(27) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27−38.
(28) Clore, G. M.; Kuszewski, J. *J. Am. Chem. Soc.* **2002**, *124*, 2866−2867.
(29) Lipsitz, R. S.; Sharma, Y.; Brooks, B. R.; Tjandra, N. *J. Am. Chem. Soc.* **2002**, *124*, 10261−10266.
(30) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 2337−2338.

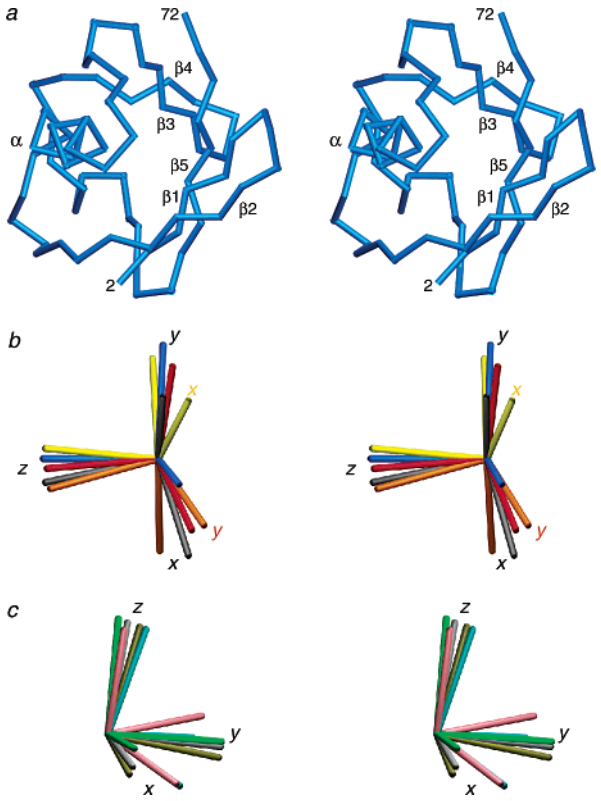

Figure 1. Orientation of the 11 alignment tensors relative to the molecular coordinates of ubiquitin. (a) Stereoview of a Cα trace of ubiquitin. (b) Axes of alignment tensors for media 1 (blue), 2 (red), 3 (gray), 4 (orange), and 5 (yellow). (c) Axes of alignment tensors for media 6 (tan), 7 (silver), 8 (green), 9 (pink), 10 (cyan), and 11 (purple). The structure and tensor orientations shown correspond to those obtained after refinement using an ensemble size of $N_e = 1$. The tensor orientations, however, for the NMR coordinates 1D3Z[12] are essentially identical. The identity of the 11 liquid crystalline media is provided in footnote a to Table 1.

the second group. The pairwise difference in orientation of the principal components of the alignment tensors within the first group ranges from 6 to 20° with a mean value of 12 ± 5°; for the second group the range extends from 2 to 20° with a mean value of 10 ± 5°. Thus, although the alignment tensors in these media display different rhombicities, the alignment tensors within each group are not entirely independent of one another. This is also evident from an examination of the normalized scalar product between pairs of alignment tensors (see Supporting Information):[23] the mean absolute values of the normalized scalar products are 0.82 ± 0.11 for pairs of alignment tensors involving media 1−5 and 0.92 ± 0.05 for pairs of alignment tensors involving media 6−11; the normalized scalar products between media 1−5 and media 6−10 range from 0.12 (between media 4 and 6) to 0.93 (between media 1 and 7). In one case, namely media 7 and 11, the overall alignment tensors differ by ≤2°, the normalized scalar product is 0.996, and the dipolar couplings in these two media are highly correlated (correlation coefficient of 0.98; see Figure 2e).

Experimental Uncertainty and Accuracy of Dipolar Coupling Data Sets. While $^1D_{NH}$ dipolar couplings can be measured with high precision, the accuracy may be considerably less depending on the exact experimental method employed. An estimate of accuracy is helpful to prevent overfitting (i.e., to ensure that improvements in agreement between observed and calculated dipolar couplings are not the result of fitting noise in the data). The simplest way to ascertain an upper bound of the accuracy of the primary data (derived from the Griesinger laboratory; ref 7) is to compare it to data collected from another laboratory under as near identical conditions as possible. Fortuitously, such data recently became available after we had

Table 1. Dipolar Coupling *R*-factors[a,b]

	X-ray[c]	NMR[d]	refined structures[e]	
	(1UBQ)	(1D3Z)	$N_e = 1$	$N_e = 2(r)$
NH Dipolar Couplings[f]				
medium 1 (62)	14.2	11.9	3.9 ± 0.1	3.5 ± 0.1
medium 2 (55)	25.3	21.1	10.5 ± 0.2	8.4 ± 0.6
medium 3 (60)	22.8	18.5	9.1 ± 0.2	6.7 ± 0.3
medium 4 (48)	18.7	16.4	8.5 ± 0.2	6.3 ± 0.6
medium 5 (55)	17.5	14.0	4.4 ± 0.2	3.9 ± 0.3
medium 6 (54)	14.8	11.4	11.9 ± 1.9	12.2 ± 1.7
medium 7 (48)	13.6	10.8	6.7 ± 0.2	6.1 ± 0.2
medium 8 (65)	15.2	13.3	13.3 ± 0.9	13.3 ± 1.2
medium 9 (56)	35.6	17.4	23.7 ± 7.3	19.5 ± 4.3
medium 10 (63)	12.5	5.3[d]	5.4 ± 0.4	4.9 ± 0.3
medium 11 (65)	10.7	6.4	5.8 ± 0.2	3.5 ± 0.1
N−C′ Dipolar Couplings[f]				
medium 10 (61)	13.0	6.4[d]	8.9 ± 0.5	9.1 ± 0.4
medium 11 (63)	12.3	7.3[d]	7.7 ± 0.4	7.1 ± 0.5
HN−C′ Dipolar Couplings[f]				
medium 10 (61)	15.6	8.8[d]	10.3 ± 0.4	9.9 ± 0.3
medium 11 (63)	17.1	11.0[d]	12.0 ± 0.3	11.2 ± 0.3
Cα−C′ Dipolar Couplings[f]				
medium 10 (58)	9.5	7.1[d]	7.4 ± 0.4	6.9 ± 0.3
medium 11 (54)	14.5	8.2[d]	8.2 ± 0.5	7.2 ± 0.4
Cα−Hα Dipolar Couplings[f]				
medium 10 (62)	17.0	6.3[d]	15.6 ± 0.8[e]	13.6 ± 0.7[e]
medium 11 (62)	16.4	4.8[d]	14.0 ± 1.0[e]	11.8 ± 0.7[e]
H_N−Hα Dipolar Couplings[f]				
medium 10 (65)	19.9	17.2	19.5 ± 0.7[e]	19.1 ± 0.7[e]

[a] The 11 liquid crystalline media are as follows: 1, CHAPSO/DLPC/CTAB (10:50:1) 5%.[7] 2, CHAPSO/DLPC/SDS (10:50:1) 5%.[7,8] 3, purple membrane fragments (2 mg/mL, 100 mM NaCl).[6,8] 4, phage pf1 5 mg/mL 50 mM NaCl.[7,8] 5, Helfrich phases (cetylpyridinumbromide/hexanol = 1:1.33, 26 mM NaBr, 5%).[7,8] 6, CHAPSO/DLPC (1:5) 5%.[7,8] 7, CHAPSO/DLPC/CTAB (10:50:1) 4%.[7,8] 8, *n*-dodecyl penta(ethylene glycol)/*n*-hexanol ($r = 0.96$).[7,8] 9, polyacrylamide gel (7%).[7,8] 10, DMPC/DHPC (3:1) 5%.[12] 11, DMPC/DHPC/CTAB (3:10:1) (5%).[12] [b] Values are in percent. The dipolar coupling *R*-factor is given by the ratio of the rms difference between observed and calculated values and the expected value of the rms difference if the vectors were randomly oriented. The latter is given by $[2D_a(4 + 3\eta^2)/5]^{1/2}$.[15] For the $^1H-^1H$ dipolar couplings, the distance between the protons is not fixed. Consequently, the denominator is given by $\{2\langle D_{obs}^2\rangle\}^{1/2}$.[15] [c] Protons were added to the X-ray coordinates using Xplor-NIH[18] which places the backbone HN proton in its standard position on a line that bisects the $C'_{i-1}-N_i-C\alpha_i$ angle. [d] The NMR structure 1D3Z[12] is the result of Cartesian coordinate refinement against all the dipolar couplings measured in media 10 and 11, with the exception of the NH dipolar coupling in medium 11 and the HN−Hα dipolar couplings in medium 10 which were not included in the structure determination. The positions of the H_N backbone amide protons are therefore determined by a combination of the covalent geometry restraints which seek to minimize any deviations from idealized covalent geometry (i.e., bond lengths, bond angles, and improper torsions) and the experimental $^1D_{NH}$ dipolar coupling and NOE data which may distort these positions very slightly. Since the values for the force constants employed for the covalent geometry restraints are much larger than those for the experimental restraints, the deviations from idealized covalent geometry in the 1D3Z NMR structure are very small (see Table 3). [e] The refined structures were refined against all dipolar couplings with the exceptions of the Cα−Hα dipolar couplings in media 10 and 11 and the H_N-Hα dipolar couplings in medium 10. Since the Cα−Hα vector and the N−H vectors are independent of each other, and since the Cα−Hα dipolar coupling data are of high quality, the cross-validated (free) Cα−Hα dipolar coupling *R*-factors provide a reliable means of assessing whether the improvement in agreement against the fitted dipolar couplings upon increasing the ensemble size is significant or simply a consequence of overfitting. In each case, the values and standard deviations reported are obtained by averaging over all 100 calculated ensembles. [f] The number of experimental dipolar couplings are listed in parentheses.

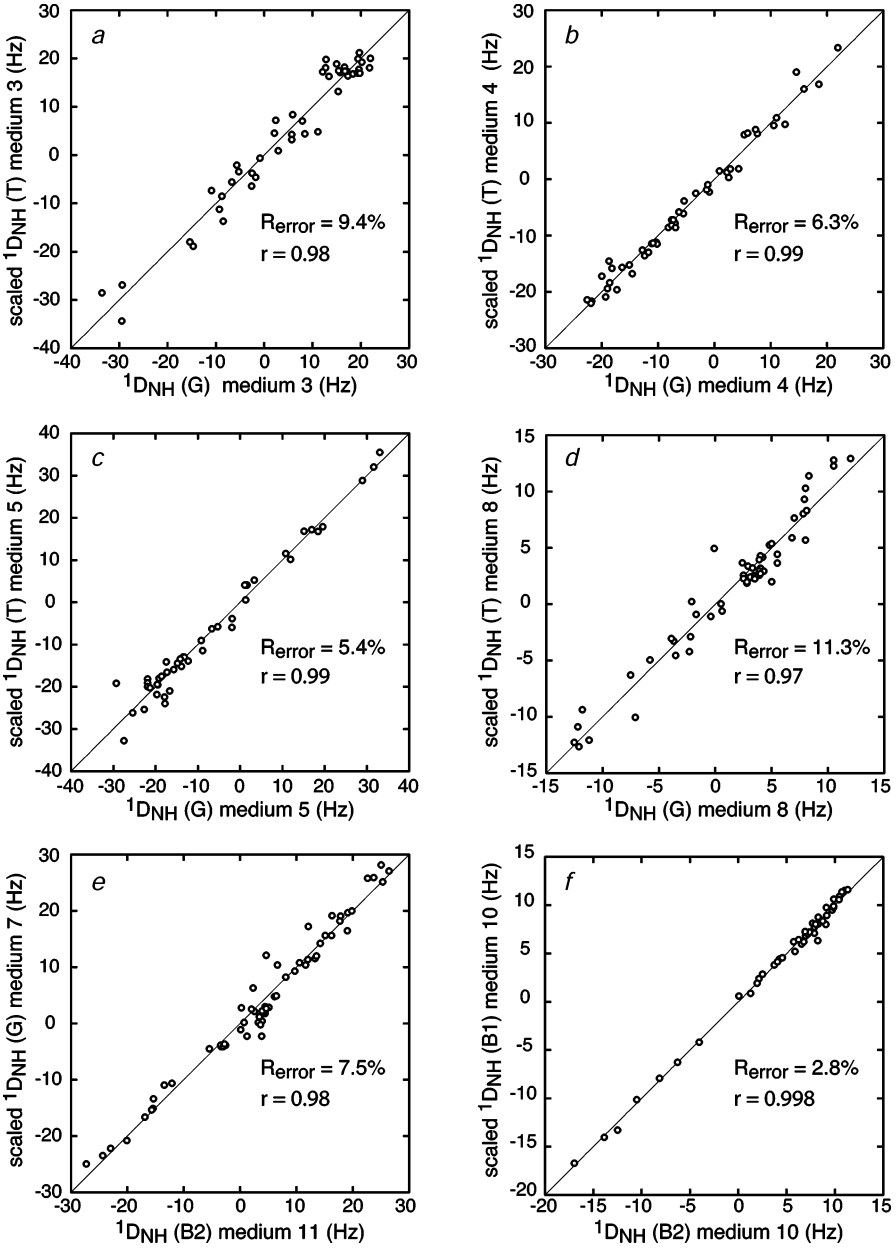

Figure 2. Assessment of accuracy in measured dipolar couplings. Correlation plots of $^1D_{NH}$ dipolar couplings measured in various media from different laboratories. Scaling of the dipolar couplings is carried out by linear regression. Accuracy is expressed as an *R*-factor error, given by $R_{err}(\%) = 100(\text{rmsd})/\sqrt{2}\zeta$ where rmsd is the root-mean-square difference between the two sets of dipolar couplings (after scaling), and $\zeta = [2D_a^2(4 + 3\eta^2)/5]^{1/2}$ is the expected rmsd if the vectors were randomly distributed[15] (D_a and η are the magnitude of the axial component of the tensor and the rhombicity, respectively; the rhombicity is defined as the ratio of the rhombic to axial component of the alignment tensor and has a maximum value of 2/3; the values of D_a^{NH} and η are obtained by nonlinear least-squares optimization using the NMR coordinates of PDB entry 1D3Z.) The letters G, T, B1, and B2 given in parentheses refer to the data from refs 7, 5c, 12, and 2a, respectively. In the case of the correlation plot between the dipolar coupling data for medium 7 (CHAPSO/DLPC/CTAB 4%) and medium 11 (DMPC/DHPC/CTAB 5%), nonlinear least-squares optimization against the NMR (1D3Z) coordinates indicates that the orientation of the two alignment tensors differ by only 1.6° and the rhombicity by only 0.1 (0.6 versus 0.5). The equivalent media in refs 6 and 7 are not necessarily absolutely identical; thus, the concentration of salt used for medium 3 (purple membrane fragments) is slightly different (100 mM NaCl in ref 6 versus 80 mM NaCl in ref 7), the phage and salt concentrations are slightly different for medium 4 (5 mg/mL pf1 and 50 mM NaCl in ref 7 versus 3.5 mg/mL pf1 and 20 mM NaCl in ref 5c), and the concentrations of cetylpyridinumbromide/hexanol and NaBr are slightly different for medium 5 (5% and 25 mM NaBr in ref 7 versus 3.4% and 75 mM NaBr in ref 5c). These differences, however, are expected to minimally impact either the orientation of the alignment tensor or the rhombicity. These slight media differences, however, do affect the value of D_a^{NH}, and this is accounted for in the comparisons by appropriate linear scaling.

completed our calculations. Correlation plots between data from the Griesinger[7] and Tolman[5c] laboratories for media 3−5 and 8 are shown in Figures 2a−d. After linear scaling to take into account differences in D_a (arising from small differences in media concentration), there is no evidence of any systematic deviation from a linear fit with a slope of 1 and a random distribution of residuals. The dipolar coupling *R*-factor error, R_{error}, and the correlation coefficient range from 5 to 11% and 0.97 to 0.99, respectively. As noted above, the dipolar coupling data recorded by the Griesinger laboratory[7] for medium 7 and by the Bax laboratory[12] for medium 11 are also highly correlated with an R_{error} value and correlation coefficient of 7.5% and 0.98, respectively. This is perhaps not too surprising because both media consist of lipid bicelles (CHAPSO/DLPC in one case

Table 2. D_a^{NH} and Rhombicity for the Different Alignment Tensors[a]

	X-ray	NMR	refined structures	
	(1UBQ)	(1D3Z)	$N_e = 1$	$N_e = 2(r)$
medium 1	28.5/0.36	28.4/0.32	29.2 ± 0.2/0.29 ± 0.01	30.0 ± 0.2/0.29 ± 0.01
medium 2	15.8/0.40	16.0/0.39	17.3 ± 0.1/0.38 ± 0.01	18.1 ± 0.3/0.39 ± 0.02
medium 3	−16.6/0.00	−17.1/0.05	−18.2 ± 0.1/0.06 ± 0.02	−19.3 ± 0.3/0.09 ± 0.02
medium 4	12.7/0.42	12.9/0.49	14.1 ± 0.1/0.40 ± 0.01	14.6 ± 0.3/0.43 ± 0.03
medium 5	20.9/0.22	20.4/0.25	22.4 ± 0.2/0.22 ± 0.01	22.8 ± 0.4/0.21 ± 0.02
medium 6	−5.9/0.22	−5.8/0.21	−5.9 ± 0.1/0.18 ± 0.06	−6.0 ± 0.1/0.21 ± 0.02
medium 7	−13.8/0.59	−13.4/0.61	−13.4 ± 0.1/0.64 ± 0.07	−13.7 ± 0.2/0.63 ± 0.01
medium 8	−7.3/0.30	−7.1/0.30	−7.0 ± 0.1/0.29 ± 0.02	−7.3 ± 0.1/0.28 ± 0.02
medium 9	2.4/0.38	−2.5/0.65	−2.6 ± 0.1/0.47 ± 0.10	−2.7 ± 0.1/0.51 ± 0.04
medium 10	−9.6/0.16	−9.8/0.16	−9.5 ± 0.1/0.16 ± 0.01	−9.8 ± 0.1/0.16 ± 0.01
medium 11	−15.4/0.49	−15.6/0.50	−15.0 ± 0.1/0.52 ± 0.01	−15.5 ± 0.1/0.52 ± 0.01
medium 10*[b]	−9.9/0.19	−9.9/0.16	−9.9 ± 0.1/0.15 ± 0.01	−10.1 ± 0.1/0.15 ± 0.01

[a] The first number listed is D_a^{NH} in Hz, and the second is the rhombicity. The values and standard deviations reported are obtained by averaging over all 100 calculated ensembles. There is a small but significant increase in the values of D_a^{NH} upon increasing the ensemble size from $N_e = 1$ to 2. This is expected as a consequence of the fact that a set of dipolar couplings arising from any distribution of the corresponding interatomic vectors will be better fit by a representation in which each dipolar coupling is represented by the average of two interatomic vectors rather than by a single vector. [b] Medium 10* (ref 2a) is nominally the same as medium 10 (ref 12), but the measurements ($^1D_{NH}$ and D_{HH}) were carried out on a different sample using a different batch of DMPC/DHPC bicelles. Consequently the values of D_a^{NH} and η are slightly different. The data collected in medium 10* were not included in the refinement calculations.

and DMPC/DHPC in the other) doped by CTAB. At the other end of the spectrum, data collected in the Bax laboratory for two different samples of media 10[2a,12] are probably as close as one can expect separate measurements to be with an R_{error} of ~3% and a correlation coefficient of 0.998. One can therefore conclude that the expected uncertainties in the $^1D_{NH}$ dipolar coupling data are consistent with dipolar coupling *R*-factors (R_{dip}) in the 5−10% range. There is one exception to this: namely, the data recorded in medium 9 where the value of D_a is only ~−2.5 Hz resulting in a small range of dipolar coupling values; assuming a typical measurement error of ~0.5−1 Hz, the expected R_{error} would be predicted to lie in the 15−30% range.

Agreement of X-ray and NMR Structures with the Measured Dipolar Couplings. Dipolar coupling *R*-factors (R_{dip}) for the X-ray (1UBQ) and NMR (1D3Z) coordinates of ubiquitin, obtained by optimization of the magnitude and orientation of the individual alignment tensors, are summarized in Table 1. For the crystal structure[12] the weighted, overall average value of R_{dip}^{NH} for the $^1D_{NH}$ dipolar couplings measured in the 11 media is 18.1%. Clearly, R_{dip}^{NH} in many of the media is larger than the expected experimental error (Table 1), reflecting uncertainties and errors in the coordinates, as well as possible differences between the solution and crystal states. The NMR coordinates, which were refined against a range of backbone dipolar couplings recorded in media 10 and 11,[12] fit the $^1D_{NH}$ dipolar couplings in media 1−8 systematically better than the crystal structure (with a weighted, overall average R_{dip}^{NH} of 15%). However, R_{dip}^{NH} for five (media 2−5 and 8) of these eight media are still significantly larger than one would expect from the anticipated experimental errors (Table 1).

The comparison of the $^1D_{NH}$ dipolar couplings with the X-ray and NMR coordinates suggests that in this instance the highly refined NMR coordinates provide a better representation of the true solution structure than the X-ray coordinates. The question then arises as to whether the discrepancies between the NMR coordinates and the $^1D_{NH}$ dipolar couplings from media 1−8 are the result of inaccuracies in the coordinates which can be removed by further refinement or whether they arise from internal motions that modulate the dipolar couplings.

Refinement against Dipolar Couplings from Multiple Media Using a Single Structure Representation. To address the question posed in the previous paragraph, we carried out a set of refinement calculations against the $^1D_{NH}$ dipolar couplings in media 1−11 and the $^1D_{NC'}$, $^2D_{HNC'}$, and $^1D_{C\alpha C'}$ couplings in media 10 and 11. The results are displayed in Tables 1−3, which summarize the calculated dipolar coupling *R*-factors, the optimized values of the magnitude (i.e., D_a^{NH} and rhombicity) of the different alignment tensors, and the agreement with other experimental NMR data, respectively. A key aspect of these calculations is that both the orientation and magnitude of the alignment tensors are refined simultaneously with the coordinates (cf. eqs 2−4). In addition to the dipolar couplings, the structures were also subject to the same interproton distance and torsion angle restraints employed to generate the 1D3Z NMR coordinates.[12,25,26] The $^1D_{C\alpha H\alpha}$ dipolar couplings in media 10 and 11, as well as the $D_{HN-H\alpha}$ dipolar couplings and various ^3J scalar backbone couplings, were not included in the refinement and provide an independent means of cross-validation. In this regard, the $^1D_{C\alpha H\alpha}$ dipolar couplings provide the most sensitive data set for cross-validation since the orientation of the Cα(i)−Hα(i) and N(i)−H(i) bond vectors are uncorrelated.[31] The $D_{HN-H\alpha}$ dipolar couplings and ^3J couplings, on the other hand, only provide a qualitative measure of cross-validation since the former are limited in their accuracy and the latter are calculated from an empirical relationship which was derived by best-fitting the observed couplings to the crystal coordinates.[17]

Refinement using a single structure representation results in very significant improvements in the R_{dip}^{NH} values for all $^1D_{NH}$ dipolar couplings with the exception of those in medium 9 (which is not unexpected because of the larger experimental errors), as well as excellent agreement with the other backbone dipolar couplings (Table 1). The weighted, overall average value of R_{dip}^{NH} is 9.3%, which represents 30 and 50% reductions relative to the NMR (1D3Z) and X-ray (1UBQ) coordinates,

(31) The covalent geometry restraints ensure that deviations from tetrahedral geometry about the Cα atom are extremely small in accordance with recent experimental results that indicate that the average angular deviations of the Cα−Ha vectors from their idealized covalent geometry are less than 1°.[11]

Table 3. Agreement with NOE-Derived Interproton Distance Restraints, ^3J Couplings, and Idealized Covalent Geometry

	X-ray	NMR	refined structures	
	(1UBQ)	(1D3Z)	$N_e = 1$	$N_e = 2$(r)
	rms from NOE-Derived Interproton Distance Restraints (1119) (Å)[a,b]			
$r_{NOE}^{(struct)}$	0.093	0.000	0.004 ± 0	0.003 ± 0.002
$r_{NOE}^{(ens)}$				0.000 ± 0
	rms from ^3J Couplings (Hz)[a,c]			
$^3J_{HNH\alpha}$ (63)	0.71	0.63	0.69 ± 0.04	0.74 ± 0.03
$^3J_{HNC\beta}$ (60)	0.31	0.26	0.34 ± 0.01	0.32 ± 0.01
$^3J_{HNC'}$ (61)	0.46	0.49	0.54 ± 0.02	0.55 ± 0.01
$^3J_{C'H\alpha}$ (65)	0.29	0.28	0.31 ± 0.01	0.32 ± 0.01
$^3J_{C'C\beta}$ (57)	0.18	0.14	0.20 ± 0.01	0.20 ± 0.01
$^3J_{C'C'}$ (56)	0.25	0.21	0.26 ± 0.01	0.24 ± 0.01
	rms from χ_1 Side-Chain Torsion Angle Restraints (35) (deg)[a,b]			
	0.10	0	0.28 ± 0.34	0.07 ± 0.11
	Deviations from Idealized Covalent Geometry			
bonds (Å)	0.017	0.006	0.004 ± 0	0.004 ± 0
angles (deg)	3.07	0.81	0.73 ± 0.02	0.75 ± 0.03
improper torsions (deg)	3.26	0.68	1.29 ± 0.06	1.30 ± 0.07

[a] The number of experimental terms are listed in parentheses. [b] The NOE-derived interproton distance restraints[25] and χ_1 side-chain torsion angle restraints[26] are included in the target function used for refinement. There are no NOE violations >0.2 Å or torsion angle violations greater than 5° for either the 1D3Z NMR coordinates or the refined structures. For the X-ray coordinates (1UBQ) there are 10 interproton distance violations greater than 0.5 Å. $r_{NOE}^{(ens)}$ and $r_{NOE}^{(struct)}$ are defined in eqs 12 and 13, respectively. [c] The ^3J couplings are *not* included in the target function for refinement and therefore serve as an independent check on the results. Note that the coefficients for the Karplus equations relating ^3J to torsion angles were derived by initially best-fitting to the X-ray (1UBQ) coordinates.[17] For the ensemble size $N_e = 2$ calculations, the reported values represent the ensemble averaged values.

Table 4. Backbone Atomic rms Differences (Å)

	ensemble size	
	$N_e = 1$	$N_e = 2$(r)
intra-ensemble		0.43 ± 0.07
mean coordinates of each ensemble		
versus overall refined mean	0.17 ± 0.04	0.12 ± 0.02
versus NMR (1D3Z)	0.52 ± 0.03	0.44 ± 0.03
versus X-ray (1UBQ)	0.58 ± 0.03	0.53 ± 0.03
average intra-ensemble rigid body rotational difference (deg)		0.3

respectively. Moreover, the R_{dip} values for all dipolar couplings included in the refinement are, to a first approximation at least, comparable to their expected experimental error. The $R_{dip}^{C\alpha H}$(free)[32] values for the cross-validated $^1D_{C\alpha H\alpha}$ dipolar couplings are better than the corresponding $R_{dip}^{C\alpha H}$ values for the X-ray coordinates (Table 1), and no significant degradation in the agreement with either the $D_{HNH\alpha}$ couplings (Table 1) or the ^3J backbone couplings (Table 3) is observed. Moreover, there is no evidence for any deviation from a slope of 1 for the correlation between observed and calculated $D_{HNH\alpha}$ dipolar couplings, using the alignment tensor derived from the $^1D_{NH}$ dipolar couplings recorded in the same medium (i.e., medium 10*; see Supporting Information). Further, both the NOE-derived interproton distance restraints and the side-chain torsion angle restraints are satisfied within experimental error (Table 4). We also note that the planarity of the peptide bond (including the position of the N−H vector in the peptide plane) is maintained close to ideality and that the deviations from planarity (maximum 5° with an average of 2.5°) are well within the range observed in high-resolution (1.75 Å or better) protein crystal structures.[22]

These results would suggest that a single model representation provides, to a first approximation at least, a rather good representation of the experimental data and that there is little need to invoke large-scale internal motions.

Refinement against Dipolar Couplings from Multiple Media Using an Ensemble Representation with $N_e = 2$. Although a single structure representation, after refinement, fits the experimental $^1D_{NH}$ dipolar coupling data well, one can ask the following question: can a multiple structure representation result in further significant improvements in the agreement between observed and calculated $^1D_{NH}$ dipolar couplings? While simple librations may result in isotropic motion of an N−H vector (i.e., wobbling in a cone), any motion that arises from changes in backbone torsion angles will result in anisotropic motion. The simplest representation of anisotropic motion is a two-structure ensemble, which corresponds to hopping between two equally populated states at a rate that is fast on the chemical shift time scale (since only a single set of resonances is observed). The overall calculated dipolar couplings for the ensemble are given by the average of the calculated dipolar couplings of the individual members of the ensemble. It is important to stress that in this formulation averaging occurs at the level of the dipolar couplings and not the coordinates; thus, the dipolar couplings computed in this manner are not the same as those computed from the average coordinates of the ensemble members which would be physically meaningless and obviously not agree with the experimental data.

To obtain insight into the physical meaning of the two-structure ensemble representation to model the amplitudes of local motions, consider the following simple example in which the N−H vectors of two residues, k and l, each exist in two distinct orientations, k_+ and k_-, l_+ and l_-, respectively. The same calculated values of the average $^1D_{NH}$ dipolar couplings for residues k and l will be obtained for two distinct pairs of structures: k_+l_+/k_-l_- and k_+l_-/k_-l_+. (Note that the other energy terms in the target function can be different for the two pairs of structures.) A single two-structure ensemble can therefore only represent a snapshot of the system. By calculating a large number of two-structure ensembles, however, the equilibrium

(32) The notation R_{dip}(free) refers to the dipolar coupling R-factors for the cross-validated (i.e., free) set of dipolar couplings that are excluded from the refinement. R_{dip}(work) refers to the dipolar coupling R-factors for the working set of dipolar couplings; that is, those dipolar couplings that are included in the target function employed for refinement.

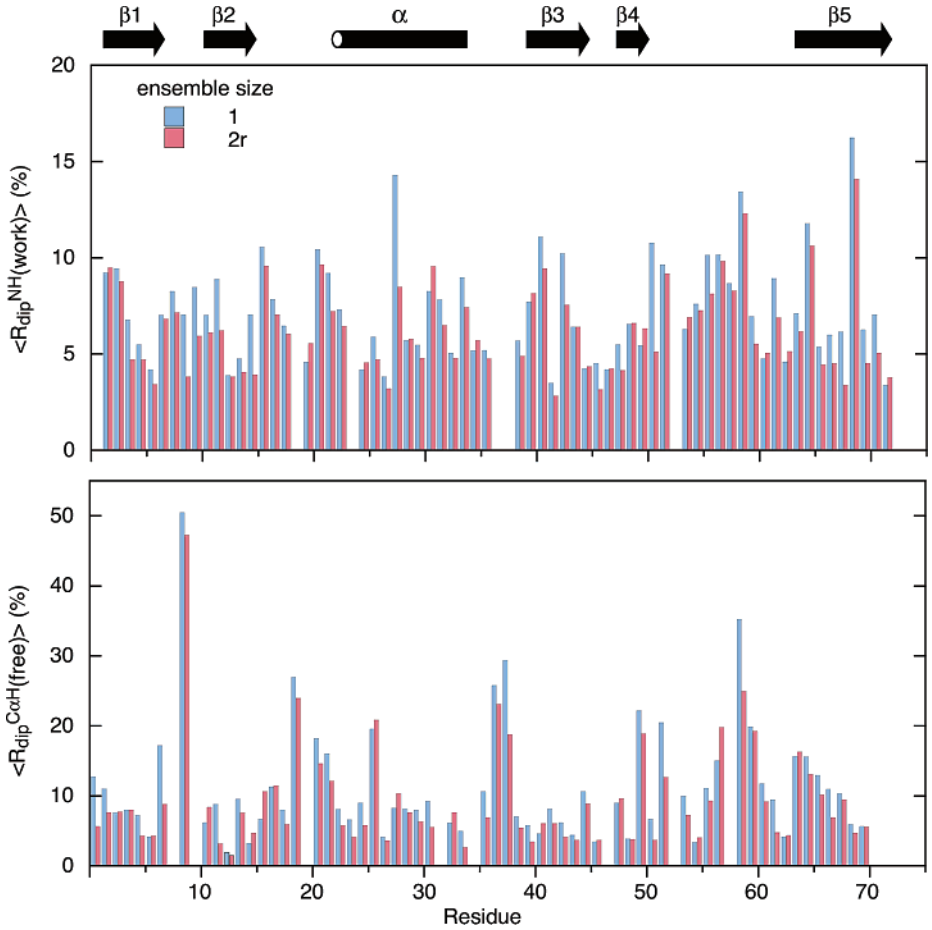

Figure 3. Agreement between observed and calculated working and cross-validated dipolar couplings as a function of residue number for structures refined with ensemble sizes of $N_e = 1$ (blue) and 2 (red). The working set of dipolar couplings comprises the NH dipolar couplings measured in all 11 media, and the agreement is expressed as a residue-based *R*-factor, $\langle R_{dip}^{NH}(\text{work})\rangle$, averaged over all media; the cross-validated (free) set of dipolar couplings comprise the CαH dipolar couplings measured in media 10 and 11, with the agreement expressed as a residue-based free *R*-factor, $\langle R_{dip}^{C\alpha H}(\text{free})\rangle$, averaged over the two media. (The angle brackets $\langle\rangle$ denote averaging over all 100 calculated ensembles.) The results for the ensemble size of $N_e = 2$ constitute those obtained from the restricted 2r set of calculations (see Methods). The location of secondary structure elements is shown at the top of the figure.

distribution between the two combinations, and hence the extent of correlation between the N−H vector orientations of residues *k* and *l*, is obtained. If the two distinct combinations are equally populated, no correlation exists; if only one combination is observed, the orientations of the two N−H vectors are fully correlated; an intermediate situation between these two extremes indicates partial correlation. In the context of a protein structure refinement, the degree to which local motions are correlated or uncorrelated is therefore obtained by examining the distribution of N−H vectors within a large collection of calculated ensembles (in this instance 100). In general, the correlations involving the majority of local motions are expected to extend only to adjacent neighbors. This is manifested by compensatory changes in backbone torsion angles such that the fold of the protein molecule remains unperturbed. This is discussed in detail below.

Calculations with an ensemble size $N_e = 2$ were carried out using a variety of conditions permitting different ranges for differences in relative atomic positions, values of D_a^{NH}, and rhombicity among members of an ensemble (see Methods section). The restricted (2r) set of calculations corresponds to the case where the shape of the molecule remains unchanged on a time scale longer than the overall rotational correlation time (τ_c), and thus the ensemble can be described by a single alignment tensor. The alignment tensor is very sensitive to molecular shape,[19] and the partially restricted (2pr) and unrestricted (2u) set of calculations correspond to cases where increasing degrees of shape changes in the molecule occur as a consequence of the internal motions. This would be the case, for example, in a situation where there was a correlated motion of the single α-helix of ubiquitin on a time scale slower than τ_c, as postulated in ref 9. The results of these various calculations all point to the same conclusions so that only the results for the minimalistic 2r set of calculations are displayed in Tables 1−3 and Figures 3−8. (Comparisons of the results of the 2pr and 2u set of calculations with those of the 2r set of calculations are provided in the Supporting Information.)

The effect of increasing the ensemble size from $N_e = 1$ to 2 is illustrated by the comparison of the overall average dipolar coupling *R*-factors for the working set of $^1D_{NH}$ dipolar couplings, R_{dip}^{NH}(work),[32] and the free cross-validated set of $^1D_{C\alpha H}$ dipolar couplings, $R_{dip}^{C\alpha H}$(free),[32] displayed as a function of residue number in Figure 3. The weighted, overall average value of R_{dip}^{NH}(work) is reduced from 9.3 to 8.0% as the ensemble size is increased from 1 to 2 (Figure 3, top panel). This represents an overall 14% improvement in agreement with the $^1D_{NH}$ dipolar couplings. Moreover, there is a statistically significant[33] reduc-

tion in R_{dip}^{NH}(work) for seven of the 11 media (Table 1). Since increasing the ensemble size from $N_e = 1$ to 2 involves an increase in the number of degrees of freedom, the possibility exists that the improvement in R_{dip}^{NH}(work) is a consequence of overfitting the data (e.g., fitting noise). This is readily tested by cross-validation. Thus, if the agreement with cross-validated observables[32] that are not correlated to those used in the refinement is improved upon increasing the ensemble size from $N_e = 1$ to 2, one can conclude unambiguously that an ensemble size of $N_e = 2$ provides a better representation of the $^1D_{NH}$ dipolar coupling data from the multiple media than the single structure representation ($N_e = 1$). This is indeed the case. There is a statistically significant[33] decrease in the free cross-validated (i.e., not included in the refinement) *R*-factors, $R_{dip}^{C\alpha H}$(free), for the $^1D_{C\alpha H}$ dipolar couplings recorded in both media 10 and 11 (Figure 3, bottom panel; Table 1). The overall reduction in $R_{dip}^{C\alpha H}$(free) upon increasing the ensemble size from $N_e = 1$ to 2 is ~15%. The agreement with the other cross-validated terms, specifically the $D_{HN-H\alpha}$ couplings and the ^3J couplings, is not affected by changing the ensemble size. This does not detract in any way from the conclusions drawn from the reduction in $R_{dip}^{C\alpha H}$(free) and largely reflects the greater uncertainties in the $D_{HN-H\alpha}$ and ^3J coupling data (in terms of either measurement error in the case of the former or in limitations in the empirical relationship between dihedral angles and ^3J couplings in the case of the latter).

We also note that there is no significant difference in the agreement between observed and calculated dipolar couplings (both free and working sets) for the restricted (2r), partially restricted (2pr), or unrestricted (2u) sets of calculations with an ensemble size of $N_e = 2$ (see Supporting Information). The weighted, overall average R_{dip}^{NH} for these three sets of calculations is 8.0, 8.0, and 7.8%, respectively. This is important because it indicates that allowing larger differences between ensemble members in terms of either atomic rms shifts or the magnitude of the individual alignment tensors does not afford any additional improvements. Thus, one can conclude that the minimalistic 2r set of calculations is sufficient to describe the experimental data.

The differences in the calculated dipolar couplings between the members of an ensemble vary from less than 1% up to ~18% in the case of the restricted 2r set of calculations (Figure 4). The bond vector angles made by equivalent N−H vectors of the two members of an ensemble exhibit quite a large degree of variation, but are less than 30° for all but nine residues (Figure 4, middle panel). Since for $N_e = 2$ the different N−H vector orientations between ensemble members is equivalent to a two-site jump model, the inter N−H bond vector angle θ can be conveniently expressed as an order parameter given by S^2(jump) $= (3\cos^2\theta + 1)/4$.[34,35] S^2(jump) has a maximum value of 1 and a minimum value of 0.25 when $\theta = 90°$. All but nine residues have values of S^2(jump) ≥ 0.8 (Figure 4, bottom panel).

The average intraensemble backbone atomic rms difference between two members of each ensemble in the 2r set of calculations is only ~0.4 Å (Table 4). Although this value is increased substantially for the 2pr (~0.7 Å) and 2u set of calculations (2.3 Å and ~1.0 Å before and after least-squares optimal superposition of the backbone N, Cα, C′ coordinates), the differences in N−H bond vector orientations and the calculated S^2(jump) order parameters for the three sets of calculations remain very similar (see Supporting Information). Hence, the conclusions regarding the variation in the differences in N−H bond vector orientations and corresponding order parameters between structure pairs within an ensemble (displayed in Figure 4 for the 2r set of calculations) are unaffected by the details of the calculations.

It is also worth noting that the backbone rms difference between the ensemble means and the overall mean (averaged over 100 calculated ensembles) is very small (<0.15 Å) and that the overall mean coordinates from the 2r, 2pr, and 2u calculations are almost identical (see Supporting Information). In addition, the pattern of atomic rms differences to the NMR (1D3Z) and X-ray (1UBQ) backbone coordinates is very similar for both the $N_e = 1$ and 2 calculations, although the magnitude of the average differences is reduced a little for the ensemble means from the $N_e = 2$ calculations relative to the structures from the $N_e = 1$ calculations (Figure 5 and Table 4).

Validity of Results for Residues with Small S^2(Jump) Order Parameters. There are three residues with especially small (<0.6) values of S^2(jump): A28, E51, and L69. (In the 2r set of calculations, these three residues have S^2(jump) values of 0.59, 0.34, and 0.31, respectively.) The question therefore arises as to whether the magnitude of the amplitudes of the anisotropic motions for the N−H bond vectors of these three residues are real or whether they are simply an artifact of experimental measurement error. To assess this, we carried out a series of additional calculations.

First, we repeated the 2r set of calculations omitting all NH dipolar coupling data pertinent to A28, E51, and L69. Not surprisingly, the S^2(jump) values for these three residues are increased to ≥0.96. In this calculation, however, the values of S^2(jump) for the NH bond vectors for the remainder of the protein are essentially unperturbed, and in addition, the reduction in the overall cross-validated $R_{dip}^{C\alpha H}$(free) relative to the $N_e = 1$ calculations is unaffected. This result is important for two reasons. First, it highlights the very local nature, in structural terms, of the anisotropic motions which is discussed in detail later on. Second, it shows unequivocally that the improvement in the agreement with the cross-validated $^1D_{C\alpha H\alpha}$ dipolar couplings is not affected by the presence or absence of $^1D_{NH}$ dipolar coupling data for these three outlying residues.

We next examined the agreement between observed and calculated $^1D_{NH}$ dipolar couplings for A28, E51, and L69 in the $N_e = 1$ calculations. In each case, the residue-specific R_{dip}^{NH} values in two of the 11 media are greater than 20%: media 2 and 8 for A28, media 3 and 9 for E51, and media 2 and 11 for E69. We therefore carried out a series of calculations omitting the $^1D_{NH}$ data for these media individually and in pairs for each of the three residues.

In the case of A28, $^1D_{NH}$ data are only available for seven of the 11 media (specifically media 2, 4, 6, and 8−11). The low S^2(jump) value (as well as the ~50% decrease in the residue-

(33) For the purposes of this paper, we consider a reduction in the dipolar coupling *R*-factor, R_{dip}, upon increasing the ensemble size from $N_e = 1$ to 2 to be statistically significant when the decrease in the mean R_{dip} values (averaged over 100 calculated ensembles) is greater than the sum of their standard deviations.

(34) Lipari G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559−4570.

(35) S^2(jump) is strictly a calculated quantity derived from the N−H vector orientations in the structure ensembles. This term should *not* be confused with the generalized order parameter ***S*** for small isotropic internal motions which is the order parameter that linearly scales the magnitude of the dipolar couplings.

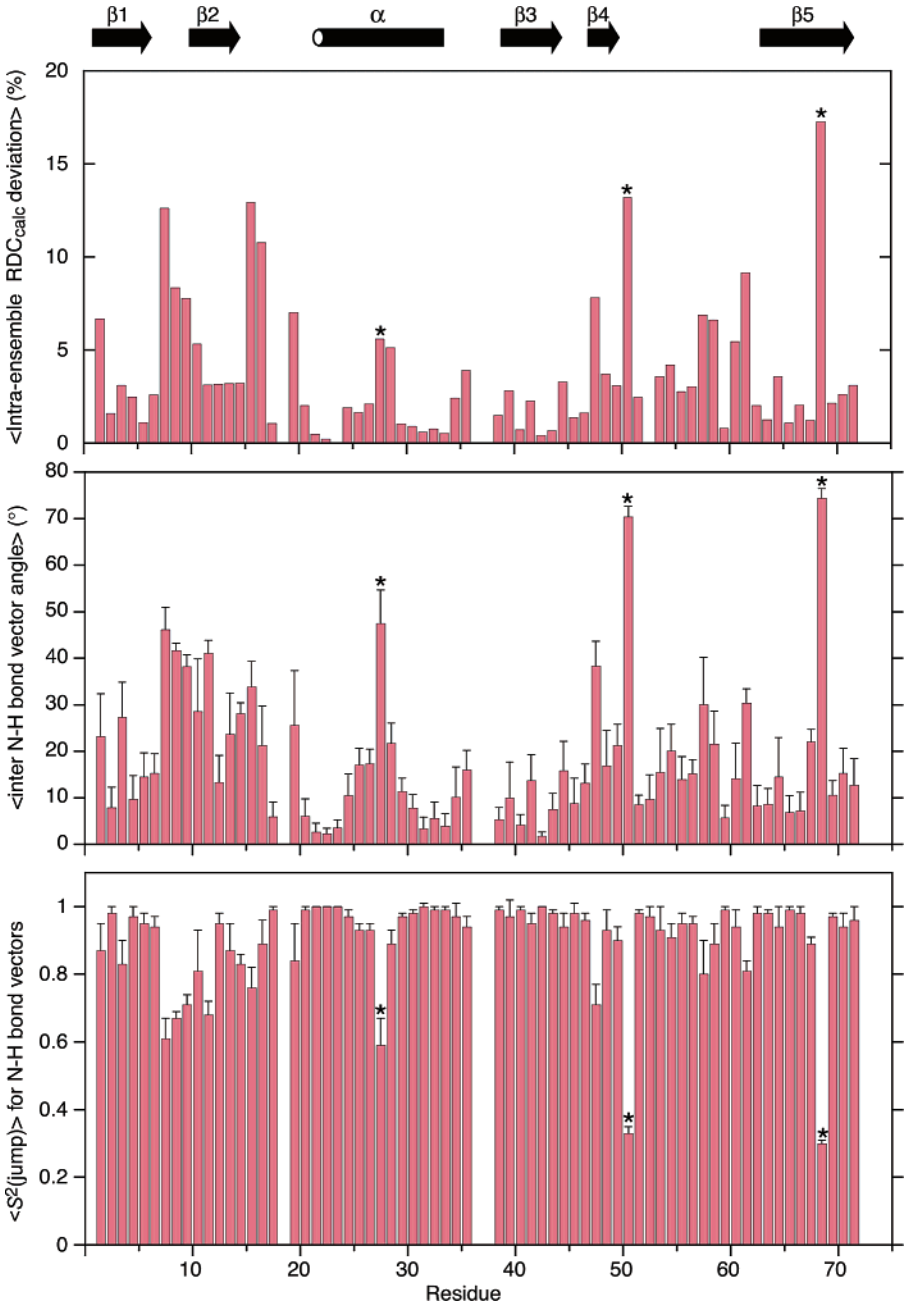

Figure 4. Structural and dynamic characteristics of the ensemble size $N_e = 2$ structures (restricted set 2r). Displayed as a function of residue number are the average intraensemble deviation in calculated $^1D_{NH}$ dipolar couplings, the average interstructure N−H bond vector angles between the members of an ensemble, and the average order parameter $\langle S^2(\text{jump})\rangle$ for the N−H bond vectors within an ensemble. The deviation in calculated dipolar couplings about the mean calculated value for each ensemble is expressed as a percentage of the total excursion in dipolar couplings given by $\langle|D_a^{NH}|\rangle(3 + 1.5\langle\eta\rangle)$. $\langle S^2(\text{jump})\rangle$ is given by $\langle(3\cos^2\theta + 1)/4\rangle$, where θ is the angle between the N−H bond vectors in the two members of an ensemble.[34] The angle brackets $\langle\rangle$ denote averaging over all 100 calculated ensembles. The error bars are the standard deviations derived from the 100 calculated ensembles. The asterisks in the figure denote the three residues (A28, E51, and K69) with especially large amplitude anisotropic N−H bond vector motions (i.e., large inter N−H bond vector angles and small S^2(jump)). As discussed in the text, it is quite likely that the motion associated with A28 is an artifact as a consequence of a potential erroneous value for the $^1D_{NH}$ dipolar coupling in medium 2; in addition, the motions associated with E51 and K69, while real, are likely to be overestimated, and a more realistic estimate of S^2(jump) is probably in the 0.5−0.6 range with a corresponding inter N−H bond vector angle of ∼45−55°.

specific R_{dip}^{NH} upon increasing the ensemble size to $N_e = 2$; cf. Figure 3, top panel) is entirely attributable to the measured value of $^1D_{NH}$ in medium 2: omitting this single dipolar coupling increases the value of S^2(jump) for A28 to 0.98, while omitting that for medium 8 has no effect on S^2(jump). This suggests the possibility that the measured dipolar coupling for A28 in medium 2 is erroneous. Unfortunately, the $^1D_{NH}$ dipolar coupling data in medium 2 are insufficiently correlated to that in the other six media to ascertain whether the reported value[7] of $^1D_{NH}$ for A28 in medium 2 is incorrect. However, we note that the average residue-based cross-validated $R_{dip}^{C\alpha H}$(free) for A28 is increased by ∼20% on increasing the ensemble size from $N_e =$ 1 to 2 (Figure 3, bottom panel), which further suggests that the N−H bond vector of A28 does not in fact exhibit any large

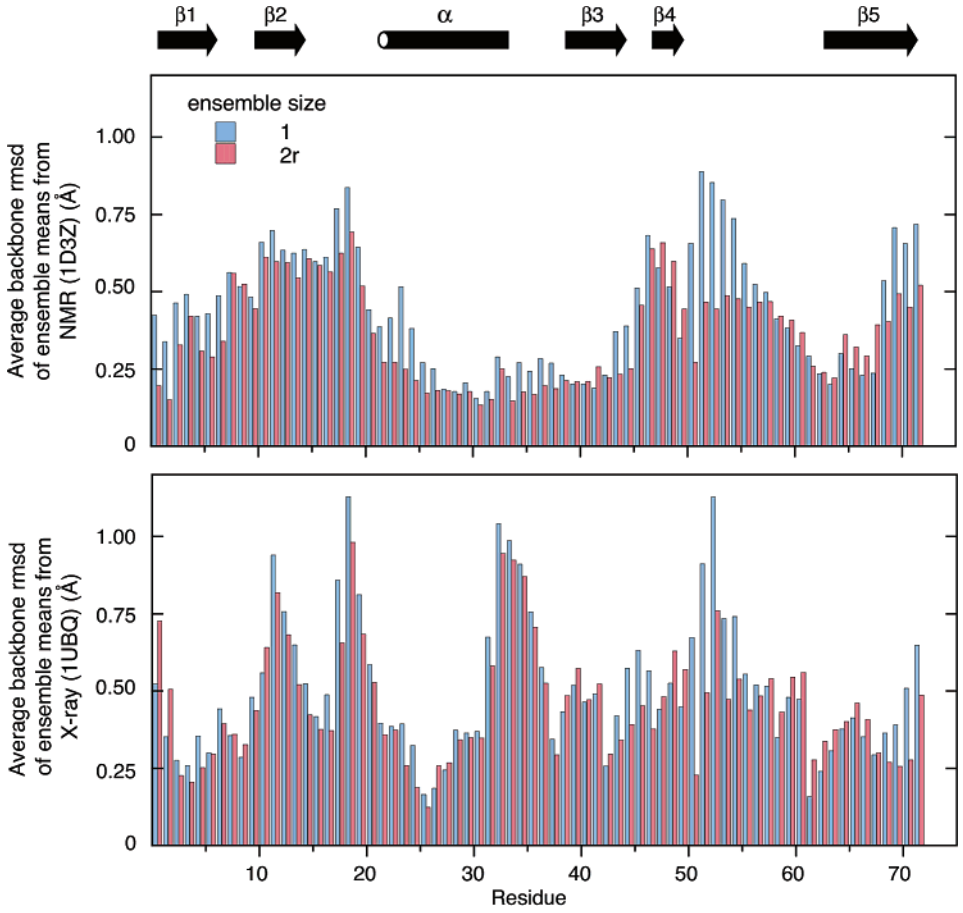

Figure 5. Comparison of the structures refined with an ensemble size of $N_e = 1$ (blue) and 2 (red) with the NMR (1D3Z; ref 12) and X-ray (1UBQ; ref 10) coordinates.

magnitude anisotropic motion. Given that A28 is located in the middle of an 11-residue α-helix (residues 23−33), it would hardly be surprising if A28 were indeed rather rigid.

For E51, the $^1D_{NH}$ data for medium 5 make a significant contribution to the magnitude of S^2(jump). If the data for medium 5 are omitted, S^2(jump) for E51 is increased from 0.34 to 0.46 and increased slightly further to 0.49 upon the additional omission of the data from medium 9. For K69, the data from both media 2 and media 11 contribute: omitting the data from either medium 2 or medium 11 increases S^2(jump) from 0.31 to 0.42 and 0.35, respectively; omitting the data from both media 2 and 11 increases S^2(jump) still further to 0.54. Thus, in contrast to the situation with A28, substantial anisotropic motion of the N−H bond vectors still remains for both E51 and K69 when the relevant $^1D_{NH}$ data for media with the largest deviations in the $N_e = 1$ calculations are omitted. This therefore suggests that all the $^1D_{NH}$ dipolar coupling data pertinent to E51 and K69 contribute to the observed result and that while the magnitude of the anisotropic motions for the N−H bond vectors of E51 and K69 may be overestimated, substantial anisotropic motions are in fact present (with an inter N−H bond vector angle of ∼45−55° between the two members of an ensemble). This is further supported by the observation that the average residue-based cross-validated $R_{dip}^{C\alpha H}$(free) values are reduced (by ∼45 and ∼20% for E51 and K69, respectively) upon increasing the ensemble size from $N_e = 1$ to 2 (Figure 3, bottom panel).

Refinement against Dipolar Couplings from Multiple Media Using an Ensemble Representation with $N_e > 2$. If an ensemble representation with $N_e = 2$ improves the agreement between calculated and observed dipolar couplings for both the working and cross-validated free sets, it is natural to ask whether further increases in ensemble size afford yet additional improvements? While increasing the ensemble size beyond $N_e = 2$ results in a small reduction in the overall average R_{dip}^{NH}(work), it has no statistically significant effect on the overall $R_{dip}^{C\alpha H}$(free) (see Supporting Information). Although increasing the ensemble size beyond $N_e = 2$ reduces the average pairwise difference in N−H bond vector orientations between members of an ensemble, S^2(jump), which in the (equally populated) multisite case is given by $\Sigma_{ij}(1/N_e)^2 P_2$ (cos θ_{ij}) (which includes $i = j$),[34] where P_2 (cos θ_{ij}) is the second-order Legendre polynomial (3 cos^2 $\theta_{ij} - 1$)/2, remains essentially unchanged (see Supporting Information). Thus, the interpretation of the data in terms of molecular motions remains the same upon increasing the ensemble size beyond $N_e = 2$. One can therefore conclude that the jump model affords an appropriate representation of anisotropic motions and that two sites are sufficient to describe the amplitude of these motions.

Structural Implications of Ensemble Averaging. From the calculations summarized above one can conclude that (a) anisotropic motion as represented by an ensemble size of $N_e = 2$ accounts for the experimental dipolar couplings better than a single model representation, (b) increasing the ensemble size

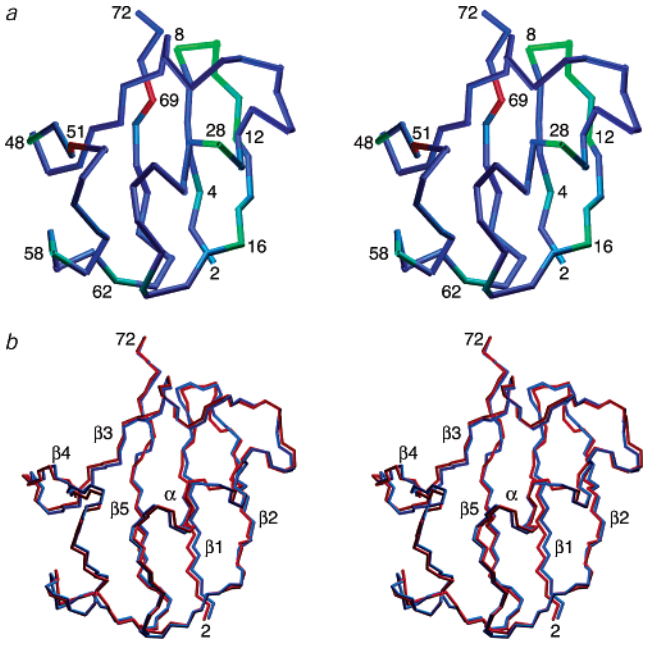

Figure 6. Stereoviews illustrating the overall structural impact of refinement using an ensemble size of $N_e = 2$. (a) Cα trace of ubiquitin color coded according to $\langle S^2(\text{jump})\rangle$ (where the angle brackets denote averaging over all 100 calculated ensembles). $\langle S^2(\text{jump})\rangle$ varies linearly through the color spectrum from red for a value of 0.3 to blue for a value of 1.0, with green corresponding to a value of ~0.6. (b) Two members (shown in red and blue) of a typical ensemble from the 2r set of calculations. The backbone coordinates have been optimally translated relative to one another but no best-fit rotation has been carried out.

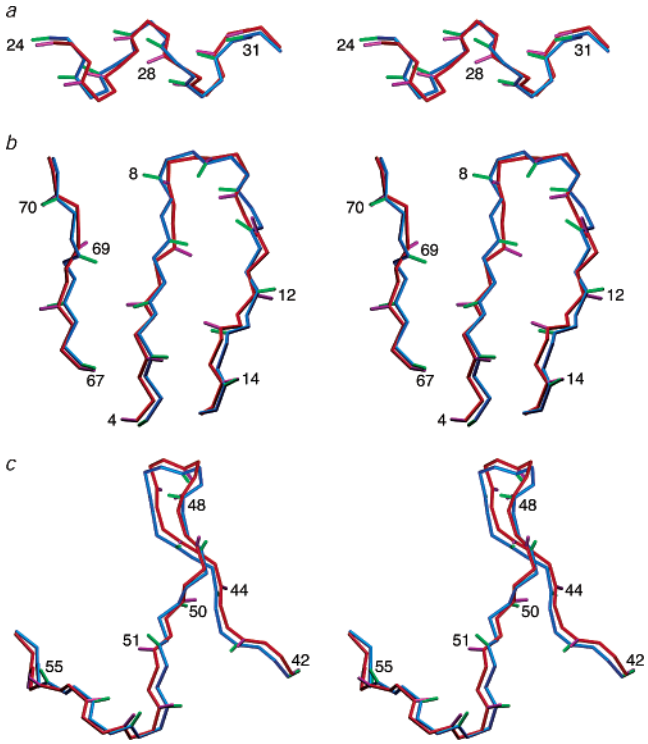

Figure 7. Stereoviews illustrating details of structural differences between two members of a typical ensemble from the 2r set of calculations for regions displaying the lowest order parameters. The backbone (N, Cα, C atoms) and N−H bonds are shown in blue and green, respectively, for one member of the ensemble, and in red and magenta, respectively, for the other.

beyond $N_e = 2$ does not improve the description of the anisotropic motions, (c) even large anisotropic $N-H_N$ motions, typified by differences in bond vector orientations >30° and S^2(jump) order parameters <0.8 can be accommodated by very small backbone atomic rms displacements of ≤0.5 Å (cf. Figure 6b), and (d) calculations in which larger atomic rms displacements are allowed to occur have minimal impact on the N−H bond vector orientations (cf. Supporting Information).

A Cα backbone trace of ubiquitin colored according to the value of S^2(jump) is shown in Figure 6a. There are six main regions characterized by residues with low (≤0.8) S^2(jump) order parameters: the loop connecting strands β1 and β2 (L8, T9, G10) and the first two residues of strand β2 (K11 and T12); the first residue (E16) of the loop connecting strand β2 to the α helix; A28 in the middle of the α-helix; the first (K48) residue of strand β4 and the residue (E51) immediately following strand β4; two residues (D58 and Q62) in the long loop connecting strands β4 and β5; and finally, L69 located close to the C-terminal end of strand β5. Thus, with only a single exception (A28 in the middle of the helix), these regions are located either in loops and turns or at the ends of secondary structure elements. Detailed views of some of these regions are shown in Figure 7, which provides a direct structural comparison of the two members of a typical ensemble of size $N_e = 2$. The key feature to notice is that the consequences of even large differences in the N−H bond vector orientations (≥50° with S^2(jump) ≤ 0.6) of a given residue between ensemble members are highly local and can be readily accommodated with minimal changes in the backbone (N, Cα, C′ atomic positions) by appropriate correlated changes in backbone ϕ and ψ torsion angles. Thus, for example, the large difference in the N−H bond vector orientations for A28 does not perturb the helix (Figure 7a), and similarly for K69, only minimally impacts strand β5 (Figure 7b). These local correlated changes in ϕ/ψ torsion angles occur within allowed regions of the Ramachandran ϕ/ψ map (see Supporting Information), and in this regard it is worth noting that 93.5 ± 2.0% of all residues lie in the most favored region of the map, with the remaining residues lying in the additionally allowed regions.[36] Concerted changes in ϕ/ψ torsion angles without significantly perturbing the backbone N, Cα, and C′ atomic positions are particularly easy to accomplish within the β region of the Ramachandran map, which spans an extensive region of torsion angle space in both the ϕ and ψ dimensions.

To assess how correlated N−H bond vector orientations are to one another, we calculated a correlation function (given by eq 20) as a function of residue for the eight residues with S^2(jump) ≤ 0.7 (Figure 8). In general, correlation between N−H bond vector orientations is limited to sequentially neighboring residues, reflecting the local nature of the motion. Thus, for example, the N−H bond vector orientations of L8 and G10 in the turn connecting strands β1 and β2 are correlated to those of the C-terminal residue of strand β1 (residue 7) and residues in strand β2 (residues 12−14). The N−H bond vector orientation of T9, on the other hand, appears to be uncorrelated to that of others. Also observed is evidence for long-range correlations between residues in close spatial proximity that are distant in the linear sequence. For example, there is some correlation between the N−H bond vector orientations of L69 (at the C-terminal end of strand β5) and those of residues in close spatial proximity located in strand β1 (residues 3 and 5), in the

(36) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(37) Fushman, D.; Tjandra, N.; Cowburn, D. *J. Am. Chem. Soc.* **1999**, *121*, 8577−8522.

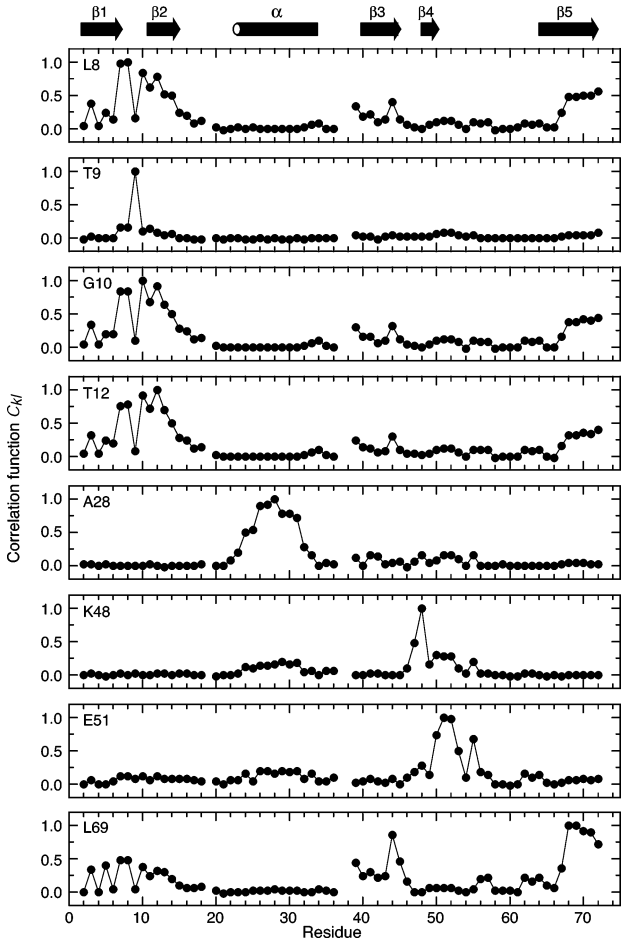

Figure 8. Correlated motions of N−H bond vectors for residues displaying the lowest order parameters, $\langle S^2(\text{jump})\rangle$, derived from the structures calculated with an ensemble size of $N_e = 2$ (2r set of calculations). The correlation function C_{kl} is plotted as a function of residue number and varies from 1.0, indicating 100% correlation, to 0, indicating 0% correlation. Since the total number of ensembles calculated is 100 and since the correlation function (eqs 19 and 20) is binary in nature, the approximate error in the correlation plots is ∼10%. The values for the order parameter $\langle S^2(\text{jump})\rangle$ in ascending order are 0.31 for L69, 0.34 for E51, 0.59 for A28, 0.61 for L8, 0.67 for T9, 0.68 for T12, 0.71 for G10, and 0.71 for K48. For all other residues, $\langle S^2(\text{jump})\rangle$ is greater than 0.75 (see Figure 4). As discussed in the text and noted in the legend to Figure 4, S^2(jump) for A28, E51, and K69 may be overestimated as a consequence of possible errors in the measured $^1D_{NH}$ dipolar coupling data for these residues in a few media. When the $^1D_{NH}$ coupling for A28 in medium 2 is removed from the calculations and S^2(jump) is increased to 0.98, not surprisingly the correlations involving the N−H vector of A28 only extend to its immediately adjacent neighbors. In the case of E51 and K69, however, S^2(jump) is only increased to ∼0.5 upon removal of the pertinent $^1D_{NH}$ dipolar coupling data (see text), and the pattern of correlations displayed in the figure is retained.

$\beta1-\beta2$ turn (residues 7, 8, and 10), at the N-terminal end of strand $\beta2$ (residue 12), and in strand $\beta3$ (residues 44 and 45). Similarly, the N−H bond vector orientations of residues L8, G10, and T12 are weakly correlated to those of the spatially close residues 68−72 at the C-terminal end of strand $\beta5$.

It is interesting to note that the correlation for A28 encompasses the full length of the helix (residues 23−33). However, these correlations involve small atomic rms displacements (Figure 7a), resulting in finely balanced compensatory changes such that the overall position and orientation of the helix remain unchanged between the two members of an ensemble (Figure 6b). It should be noted, however, that when the $^1D_{NH}$ coupling for A28 in medium 2 is removed from the calculations and S^2(jump) is increased to 0.98, not surprisingly the correlations involving the N−H vector of A28 only extend to its immediately adjacent neighbors.

The anisotropic motions probed by the dipolar coupling data can potentially extend from the picosecond to the millisecond time scale. In this regard, it is interesting to note that of the five regions with low S^2(jump) values, one has been clearly identifed by ^{15}N relaxation measurements as mobile:[36] residues in the turn between stands $\beta1$ and $\beta2$ are characterized by model free generalized order parameters $\boldsymbol{S}^2$ less than 0.8 with local correlation times in the 40−60 ps range.

Conclusion

In this paper, we have shown that the experimental $^1D_{NH}$ dipolar coupling data recorded on ubiquitin in 11 alignment media, as well as the $^1D_{NC'}$, $^2D_{HNC'}$, and $^1D_{C\alpha C'}$ recorded in two alignment media can be accounted for, at approximately the level of uncertainty in the experimental measurement, by a single model representation using restrained refinement of the coordinates and optimization of both the magnitudes and orientations of the alignment tensors. Extension to a two-member ensemble representation not only improves the agreement between observed and calculated dipolar couplings included in the refinement (i.e., the working set), but also results in a significant improvement in the agreement for the cross-validated $^1D_{C\alpha H\alpha}$ dipolar couplings which were not included in the refinement. In this representation, which provides the simplest description of anisotropic motions, the overall calculated dipolar couplings for an ensemble are given by the average of the calculated dipolar couplings for the individual members of the ensemble. This suggests that a small degree of anisotropic motion may contribute to the measured dipolar couplings. In general, the differences in N−H bond vector orientations between ensemble members is small, with a S^2(jump) order parameter ≥ 0.8. However, some residues do exhibit substantial anisotropic motion with S^2(jump) order parameters in the 0.3−0.8 range. Nevertheless, the impact of these motions on the backbone structure is minimal and, as a consequence of compensatory changes in ϕ and ψ backbone torsion angles, can be readily accommodated by very small backbone atomic rms shifts (<0.5 Å) without introducing any distortions in covalent geometry. One can therefore conclude that, for most practical applications, refinement of NMR structures against dipolar couplings using a single model representation is justified and will not adversely impact the accuracy of the resulting coordinates to any significant extent.

As in any analysis of this nature, one can only ascertain the minimum amplitude of the backbone motions required to account for the dipolar coupling data. Thus, it is clear that the residual dipolar coupling data in the 11 alignment media can be accounted for by much smaller amplitude motions than those previously proposed.[7] This, however, does not mean to say that the data cannot also be fitted to models with large-scale concerted motions. Rather, applying Occam's razor would suggest that the current experimental data provide no basis for invoking such large-scale concerted motions.

Finally, we note that the methodology of ensemble refinement introduced here provides a simple approach for detecting and

interpreting anisotropic motions (including large-scale motions) derived from residual dipolar coupling (or other appropriate) data.

Acknowledgment. We are very grateful to Attila Szabo for extensive and stimulating discussions. We also thank Ad Bax for many useful discussions and C. Griesinger, W. Peti, and A. Bax for making their residual dipolar coupling data available to us. This study utilized the high performance computational capabilities of the Biowulf/LoBoS3 cluster at the National Institutes of Health. This work was in part supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C.).

Supporting Information Available: Additional data, including alignment media, pairwise normalized scalar products of alignment tensors for ubiquitin, and additional data on refinement calculations for ensemble sizes $N_e = 2$, 4, and 8 (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA0386804