A Multifaceted Approach to the Interpretation of NMR Order Parameters: A Case Study of a Dynamic α-Helix†

Eric Johnson, Scott A. Showalter, and Rafael Brüschweiler*

Department of Chemistry and Biochemistry and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306

Received: November 24, 2007; In Final Form: February 4, 2008

The model-free approach (Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559; 4559−4570) is the standard method for the analysis of NMR relaxation data from proteins. The method assigns to each interaction vector a generalized order parameter S^2, which is a measure of the spatial restriction that the vector experiences in a molecular reference frame. A variety of techniques for the interpretation of S^2 values and their underlying assumptions are tested here for a dynamic α-helix represented by an ensemble of peptide structures that satisfies a set of predefined constraints. A self-consistent picture is developed that characterizes the influence of major determinants, including backbone dihedral angle fluctuations and their correlations, separability of internal and overall motion, the preservation of hydrogen bonds, restricted end-to-end distance fluctuations, locally anisotropic dynamics, and local contacts between interaction vector atoms and their environment. Many of the features parallel experimental NMR and computational observations of helices in proteins. The understanding gained from this model system is expected to contribute toward the ever more detailed interpretation of experimental order parameter profiles.

Introduction

NMR spin relaxation experiments of proteins provide uniquely detailed and quantitative information about conformational dynamics on the pico- to nanosecond time scale.[1−3] Over the past two decades, sophisticated spin relaxation experiments of ^{15}N, ^{13}C, or ^{2}H labeled proteins have become available for the measurement of longitudinal T_1, transverse T_2, the heteronuclear NOE, and other relaxation pathways. Interpretation of these parameters has become intimately connected to the model-free formalism by Lipari and Szabo.[4,5] The power behind the model-free approach, which is also referred to as Lipari−Szabo approach, is that it separates (using minimal assumptions) spatial dynamics from time scale effects by means of a generalized order parameter S^2, an effective internal correlation time τ_{int}, and an overall tumbling correlation time τ_c. S^2 is a measure of the spatial restriction of the reorientational motion of the dominant spin-interaction vector in a molecular reference frame, such as the ^{15}N-^{1}H bond vector in the case of backbone ^{15}N relaxation data.[6] More recently, it has become possible to determine S^2 values from residual dipolar couplings measured in multiple alignment media reporting on pico- to millisecond time scale motions.[7,8]

The Lipari−Szabo approach is widely used and is generally the first step in the analysis of experimental NMR relaxation data. In a recent survey of the literature, it was found that nearly 90% of the published NMR backbone relaxation dynamics studies use a model-free analysis of the data.[9] The appeal of the model-free approach lies in its conceptual simplicity that largely avoids the risk of overinterpretation of the experimental data. In fact, the model-free formalism was introduced (and hence its name) to overcome the arbitrary character of certain analytical motional models for the interpretation of relaxation data. These features explain the unabated popularity of the approach 25 years after its inception.

In a more general sense, a model-free analysis is the beginning of a dynamic description of a protein rather than the end. As a single real-valued quantity, the S^2 order parameter does not provide a complete description of the underlying reorientational motional distribution. This is not a deficiency of the model-free approach per se but rather reflects the actual dynamics information that relaxation data provide for a given nucleus without further interpretation involving additional assumptions. The interpretation of multiple order parameters (or relaxation parameters) for a whole protein or protein fragment, such as an α-helix, within a self-consistent structural/biophysical framework requires a more concrete description. For instance, a molecular fragment in a protein, such as an aromatic ring or a peptide plane, reorients anisotropically relative to its surroundings, which requires the simultaneous interpretation of order parameters from multiple interaction vectors together with their relative orientations. Also, the effect of correlated motions between interaction vectors belonging to different residues, for example across a loop region, needs motional models that go beyond a standard model-free description.

To meet these demands, several methods have been developed that enhance the information provided by individual order parameters or by complete order parameter profiles. Such approaches include molecular dynamics (MD) computer simulations,[10−18] normal-mode analysis,[19,20] quasi-harmonic analysis,[21−25] elastic network models,[26−28] local contact models,[27,29,30] MD-derived analytical models,[31,32] and thermodynamic models.[33−37] Some of these approaches can also be applied directly to the interpretation of spectral densities extracted from the relaxation data[38−42] or to the raw relaxation data themselves.[18]

† Part of the "Attila Szabo Festschrift".

* To whom correspondence should be addressed. Tel.: 850-644-1768. Fax: 850-644-8281. E-mail: bruschweiler@magnet.fsu.edu.

10.1021/jp711160t CCC: $40.75
Published on Web 04/01/2008 PAGE EST: 7.6

The generalized order parameter S^2 of an interaction vector is defined by[4]

$$S^2 = \frac{4\pi}{5} \sum_{m=-2}^{2} \langle Y_{2m}(\theta,\varphi)\rangle\langle Y_{2m}^{*}(\theta,\varphi)\rangle \qquad (1)$$

where the polar angles (θ,j) define the direction of the vector in a molecular reference frame and $Y_{2m}(\theta,\varphi)$ are the normalized spherical harmonics of rank 2. The angular brackets in eq 1 indicate that S^2 is defined either as a time average or over an ensemble of snapshots. Thus, experimental order parameters directly reflect the underlying conformational ensemble. It is important to note that the model-free method assumes that overall and internal motions are statistically independent. This assumption is difficult to test both experimentally and computationally. Alternative approaches to the model-free method have been developed that specifically treat correlations between overall and internal motions.[43−45] Extended MD simulations are also increasingly common, which will facilitate future tests of the assumptions underlying the model-free approach.

MD simulations offer several advantages, including the explicit consideration of time scale effects.[18] The relative complexity of the MD model, however, makes it difficult to relate order parameter profiles to specific constraints imposed on secondary structural elements. To establish relationships between S^2 values and other structural and dynamic descriptors, we analyze here finite conformational ensembles of a model system. We focus on the backbone dynamics of an α-helical model peptide consisting of 12 amino acids. A database analysis of experimental order parameters shows that the backbone ^{15}N-^{1}H S^2 profiles of helices are nonuniform exhibiting characteristic trends especially toward the ends of the helix.[46] Here, various helical ensembles are generated by systematically modifying ϕ/ψ dihedral angles both with respect to fluctuation amplitudes and their mutual correlations. From these relationships, one can establish guidelines for solving the inverse problem, which is the deduction of dynamic information about these internal degrees of freedom from experimental order parameters. The primary advantage of this approach is that it relies upon minimal assumptions, providing an unobstructed view of the underlying dynamics.

Order parameters are calculated for this system using five different methods. In addition to the standard definition of eq 1, these methods include: (1) the isotropic reorientational eigenmode dynamics method (iRED), (2) a first-order expansion in local variances and covariances, (3) the three-dimensional (3D) Gaussian axial fluctuation (GAF) method, and (4) the local contact model. Each of these methods illuminates a different aspect of order parameters and how they reflect their dynamic environment.

This work focuses on the effects of anisotropic motions, correlated motions, and the presence of short-range versus long-range motions on order parameters.[47] Short-range motion modulates the local dihedral angles in the vicinity of an interaction vector, such as the ϕ_i/ψ_{i-1} dihedral angle dynamics in the case of a N_i-H_i vector. Long-range motion, on the other hand, reorients an interaction vector via dihedral angles that are further apart, as in the case of a hinge bending motion of a whole protein fragment or fraying at the end of a polypeptide chain. It is shown how the combined use of multiple methods of analysis can provide an integrated description of these various factors.

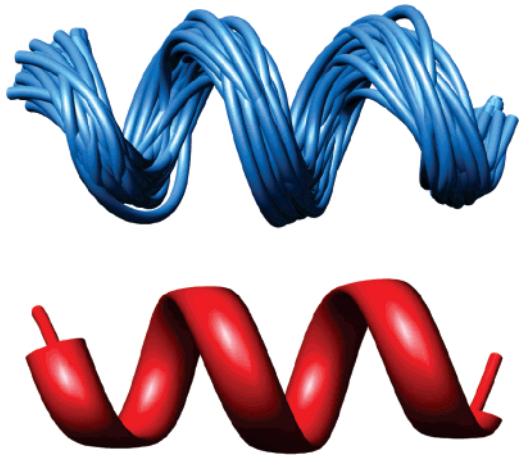

Figure 1. Helical ensemble (upper panel) consisting of 50 structures for the case in which $\sigma = 10°$ and $r = -0.5$, which is representative for the average reference helix (lower panel).

Computational Methods

A 12 residue peptide was constructed using the CHARMM22 topology file.[48] The local backbone geometry of the Leucine residue was selected for all residue subunits. A reference α-helix was generated by setting the average backbone dihedral angles to $\phi^{ref} = -57°$, $\psi^{ref} = -47°$. Ensembles were generated by randomly varying all ϕ and ψ angles according to a Gaussian probability distribution centered around the reference angles with a standard deviation σ in all angles. In addition, the distributions were skewed in such a way that they adopt a predefined linear correlation coefficient r from 0 to −1 between angle ϕ of residue i and angle ψ of residue $(i - 1)$. This procedure requires that one specify σ, r, and a tilt angle that transforms the (ϕ,ψ) distribution to a set of normal coordinates. The details of this procedure are outlined by eqs A1−A4 in the Appendix of Bremi et al.[49] A representative ensemble of 50 structures is provided in Figure 1 for the case in which $\sigma = 10°$ and $r = -0.5$. The reference helix is represented below the ensemble for comparison.

Order parameters are calculated using eq 1 after all conformers were superimposed (by translation and rotation) with respect to the reference helix. As a second method, the isotropic reorientational eigenmode dynamics (iRED) method is applied.[50,51] In this approach, where the conformers do not need to be superimposed, the isotropically averaged covariance matrix **M** is computed with elements

$$M_{ij} = \frac{1}{2}\langle 3(\mathbf{e}_i\mathbf{e}_j)^2 - 1\rangle \qquad (2)$$

where $\mathbf{e}_i$ and $\mathbf{e}_j$ are the normalized interaction vectors i and j that belong to the same conformer, and the angular brackets indicate averaging over all conformers of the ensemble. Ninety different backbone vectors are included in the analysis, corresponding to the 11 N−H, 11 N−C$^\alpha$, 11 N−C′, 11 C$^\alpha$−C′, 11 C′−O, 12 C$^\alpha$−H$^\alpha$, 12 C$^\alpha$−C$^\beta$ bond vectors as well as the 11 cross-products of the N−H and N−C$^\alpha$ vectors. The 11 cross-product vectors are included in addition to the in-plane N−H, N−C$^\alpha$, N−C′, C$^\alpha$−C′, and C′−O vectors to comprehensively probe reorientational motion of the peptide planes. Next, a principal component analysis is performed by solving the eigenvalue problem $\mathbf{M}|m\rangle = \lambda_m|m\rangle$ where $m = 1, ..., N$, and N is the number of vectors. λ_m are the eigenvalues sorted according to size (first eigenvalue is largest), and $|m\rangle$ are the associated eigenvectors. The S_j^2 value for vector j was then calculated as[50]

$$S_j^2 = 1 - \sum_{m=6}^{N} \lambda_m |\langle m|j\rangle|^2 \quad (3)$$

where the sum extends over the eigenvectors $|m\rangle$ that belong to the N − 5 smallest eigenvalues λ_m, which correspond to internal motions. There is no need to align the conformers because the covariance matrix **M** contains only inner products and therefore is rotationally invariant. In fact, one of the main advantages of the iRED method is that it makes no assumptions regarding the separability between internal and overall tumbling motions. A separability index g follows naturally from the analysis and was calculated as

$$g = \sum_{m=1}^{N} \lambda_m / \sum_{m=6}^{N} \lambda_m \quad (4)$$

A large g value reflects good separability between internal and overall motions, which is typically found for globular proteins. In contrast, small g values are observed in disordered systems where internal and overall tumbling motions are not independent. Other issues related to separability have been discussed by Meirovitch et al.[43]

The 3D GAF model provides an analytical description of anisotropic peptide plane motion.[31,32] The model represents the internal dynamics of the peptide plane in terms of independent Gaussian axial fluctuations about three orthogonal axes α, β, and γ. Reorientational motion can then be visualized as ellipsoids with principal axes parallel to α, β, and γ and lengths that are proportional to the fluctuation amplitudes σ_α, σ_β, and σ_γ. In this respect, the 3D GAF model acts as the NMR analogue to the anisotropic temperature factor in X-ray crystallography, although the latter is also sensitive to translational motion. The 3D GAF model has recently been applied to the interpretation of residual dipolar couplings[52] as well as $^{13}C'-^{13}C^\alpha$ cross-relaxation rate constants.[53]

The local contact model was originally developed to predict the order parameter directly from a 3D protein structure.[29] In the case of the N−H bond vector for residue i, the calculation is based on the close contacts experienced by the amide H and the carbonyl O of the preceding residue with the surrounding heavy atoms k

$$S_i^2 = \tanh[0.8 \sum_k \exp(-d_{i-1,k}^{O}) + 0.8\exp(-d_{i,k}^{H})] - 0.1 \quad (5)$$

where $d_{i-1,k}^{O}$ and $d_{i,k}^{H}$ denote the distances (in angstroms) between heavy atom k and the carbonyl O of amino acid $(i-1)$ and the amide H of amino acid i, respectively, shortened by 1.2 Å.

Results

Order Parameter Profiles. Ensembles were generated for the 12 residue peptide first with a standard deviation in the ϕ/ψ dihedral angles of $\sigma = 10°$. Dihedral angle fluctuations of this magnitude are consistent with a recent MD simulation of ubiquitin at 300 K.[18] In that simulation, of which the first 10 nanoseconds are analyzed in the present study, the standard deviation in the dihedral angles of the central helix (residues 23−34) varies between 7.7 and 19.7° for ϕ and 7.3 and 15.4° for ψ. S_{iRED}^2 values for the 8 different bond vector types are plotted as a function of residue number in Figure 2 for four different ensembles, each consisting of 10 000 conformers and $\sigma = 10°$, but with different linear correlation coefficients r between the dihedral angles ϕ_i and ψ_{i-1}. It should be noted that these ensembles include neither local librational motions[54] nor zero-point vibrational motions,[19,20] both of which are present in the experiment and which will reduce the S^2 values. Therefore, the computed S^2 values presented here tend to be higher than their experimental counterparts.

In the uncorrelated case ($r = 0$), all of the backbone bond vectors exhibit similar S^2 profiles with gradually reduced S^2 values toward both termini; a maximum S^2 value of around 0.9 occurs near the middle of the sequence, while a minimum value of around 0.7 is observed at the ends (Figure 2a). This fraying effect is accompanied by a relatively low separability index ($g = 5.5$) between overall and internal rotation, which suggests that the capability to distinguish between internal motion and overall orientation is limited. In contrast, when anticorrelations in ϕ_i/ψ_{i-1} dihedral angles are introduced the fraying decreases and the separability increases (Figure 2b−d). Anticorrelated variations in ϕ_i/ψ_{i-1} are commonly referred to as crankshaft motions,[55−60] which are structure-conserving in the sense that they minimize distortions relative to the reference helix. In the limiting case where ϕ_i/ψ_{i-1} are perfectly anticorrelated ($r = -1.0$; Figure 2d), there is little or no fraying and the separability index is notably high ($g = 31.0$). In this case, the S^2 values fall into two clusters that reorient differently as a result of ϕ/ψ fluctuations. One cluster exhibits S^2 values near 1.0 and includes all four bond vectors that are attached to a C^α atom ($N-C^\alpha$, $C^\alpha-C'$, $C^\alpha-H^\alpha$, and $C^\alpha-C^\beta$). Despite variations in ϕ/ψ ($\sigma = 10°$), the anticorrelated nature of the fluctuations results in very little reorientational motion of these bond vectors. The second cluster exhibits order parameters well below 1.0 and includes all of the bond vectors that are not attached to a C^α atom (N−H, N−C′, C′−O, and (N−H) × ($N-C^\alpha$)). Among this second class of bond vectors, the order parameters are around 0.94 for the N−H, N−C′, and C′−O vectors and 0.93 for the (N−H) × ($N-C^\alpha$) vectors. These observations are consistent with crankshaft motions that keep the C^α positions fixed with respect to both translation and reorientation and allow for limited fluctuations of the peptide plane.

The limiting cases represented in Figure 2a,d provide useful reference points, but they are not observed in MD simulations, which typically yield r values of around −0.5 for α-helices. Figure 2b,c illustrate the S^2 profiles for partial correlations of $r = -0.5$ and −0.75. Both ensembles exhibit relatively high separability between overall and internal rotation ($g = 8.8$ and $g = 13.4$, respectively). Some fraying is observed even at $r = -0.75$, although it is significantly attenuated relative to the uncorrelated case.

Experimental observations indicate that proximity to helix termini does in fact correlate with lower average order parameters, but the magnitude of the effect is small. Goodman et al. compiled a database of 1855 order parameters from 20 independent NMR relaxation studies.[46] They found that the change in S^2 per residue is around 0.003 and 0.01 for the six residues near the N− and C− termini, respectively, of α-helices. One likely cause is the fact that the N−H groups of the first four residues and the C′−O groups of the last four residues are not hydrogen bonded, leading to increased flexibility.

As mentioned above, the iRED method allows calculation of S^2 values without requiring superposition of the conformers. For the $r = -0.5$ ensemble, order parameters for the N−H bond vectors were determined using both the iRED method as well as using the standard relationship from eq 1 that requires explicit superposition. Structures were superimposed on the reference helix by a least-squares method using the N, C^α, C′, and O atoms. Figure 3 demonstrates that the S^2 values obtained by eq

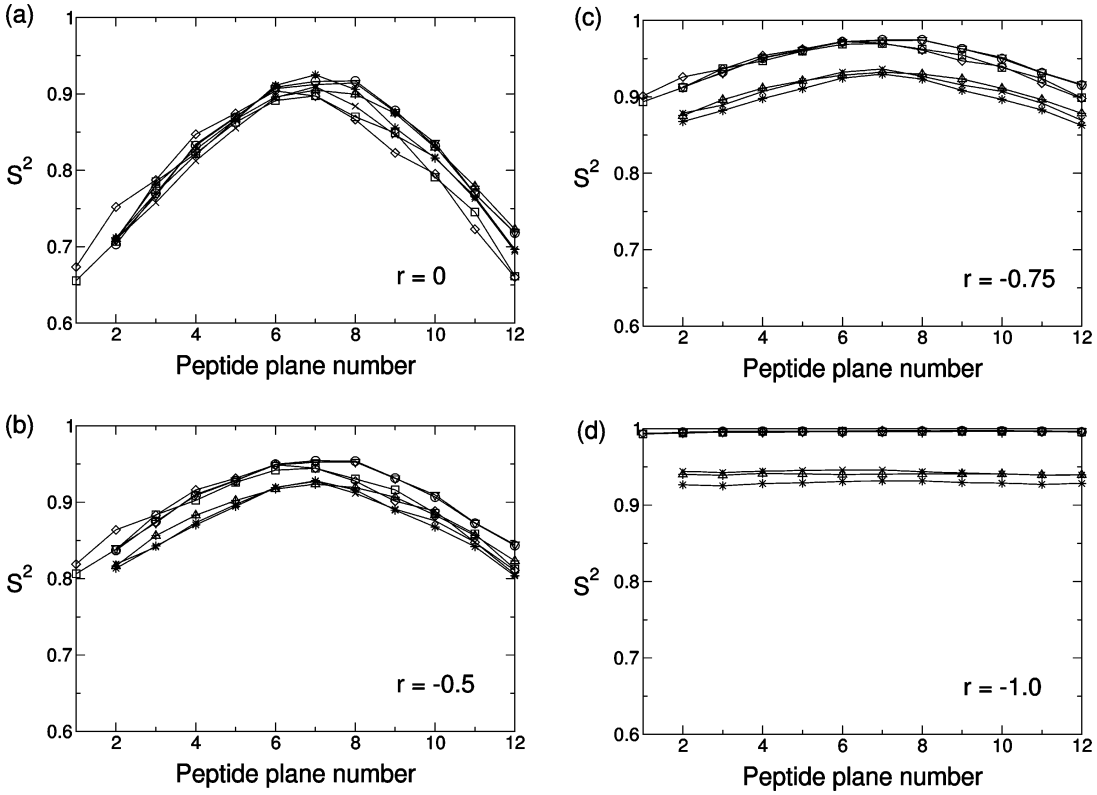

Figure 2. The iRED order parameter S^2 plotted as a function of the peptide plane number for ϕ_i/ψ_{i-1} correlations $r = 0$ (a), $r = -0.5$ (b), $r = -0.75$ (c), and $r = -1$ (d). The following bond vectors are represented by the symbols in parentheses: N−H (+), N−C$^\alpha$ (circle), N−C′ (x), C$^\alpha$−C′ (down triangle), C′−O (up triangle), C$^\alpha$−H$^\alpha$ (square), C$^\alpha$−C$^\beta$ (diamond), and (N−H) × (N−C$^\alpha$) (*). The peptide planes are numbered by the number of the amino acid residue that contributes the N atom.

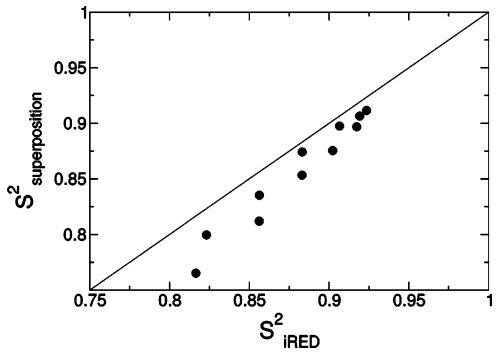

Figure 3. A correlation plot between the S^2 value calculated using eq 1 that requires superposition and the iRED S^2 value for the N−H vector in the $r = -0.5$ case.

1 are slightly lower than the iRED order parameters. The same trend, although on a smaller scale, is observed in the MD simulation of ubiquitin. In both cases, the effect is most pronounced in the flexible termini. The MD simulation yields a higher separability index than the $r = -0.5$ ensemble ($g =$ 9.6 and 8.8, respectively), which contributes to the fact that the discrepancy is smaller for the MD ensemble. The $r = -0.5$ ensemble reflects the transition toward the situation where overall and internal motions are no longer separable. The principal component analysis that underlies the iRED approach attributes the maximal amount of motion to overall reorientation (5 largest modes) so that the remaining internal modes yield a higher S^2 value.

Correlations Introduced by Structural Constraints. In the previous ensembles, the introduction of ϕ/ψ correlations generates ensembles with little fraying and relatively high separability between overall and internal rotation. Alternatively, a much larger ensemble with 10^6 conformers was generated with $\sigma =$ 10° and $r = 0$ that was trimmed according to structural criteria, such as backbone−backbone hydrogen bonds or the end-to-end distance. These criteria were intended to mimic the constraints imposed on the helix by the surrounding parts of the protein. The ensemble was first trimmed by selecting only those conformers in which all of the O and H atoms that participate in native (i, $i + 4$) hydrogen bonds are within a distance of 1.6 and 2.6 Å. This criterion selects 7639 structures from the original ensemble. The relatively small number of structures that were selected reveals that, in the absence of any dihedral angle correlations, ϕ/ψ fluctuations of this magnitude ($\sigma = 10°$) generally do not permit the preservation of all hydrogen bonds. Moreover, the sculpted ensemble that fulfills the hydrogen-bonding criterion exhibits a small but statistically significant correlation between ϕ_i and ψ_{i-1} ($r = -0.28$). As a result of the trimming procedure, the average standard deviations in ϕ and ψ also decrease slightly to 8.7 and 9.4°, respectively.

In a next step, the ensemble was further trimmed by selecting only those structures whose end-to-end distance is ± 1 Å from that of the reference helix. This second step selects 5848 of the 7639 structures that met the initial hydrogen-bonding criterion. The standard deviations in ϕ and ψ are essentially unchanged as a result of this second trimming step ($\sigma =$ 8.6 and 9.2°), whereas the amount of anticorrelation increases to a value of $r = -0.33$. The S^2_{iRED} profile for this double trimmed ensemble is shown in Figure 4. As a consequence of the trimming procedures, the fraying effect is relatively small and comparable to what is observed for the case in which the ϕ_i/ψ_{i-1} correlation was explicitly set to −0.75 (Figure 2c) with almost identical separability indices ($g = 13.2$ for the double trimmed ensemble versus 13.4 for the $r = -0.75$ ensemble). The trimming procedure, therefore, produces a reasonable order parameter profile, although experimental profiles for helices in intact

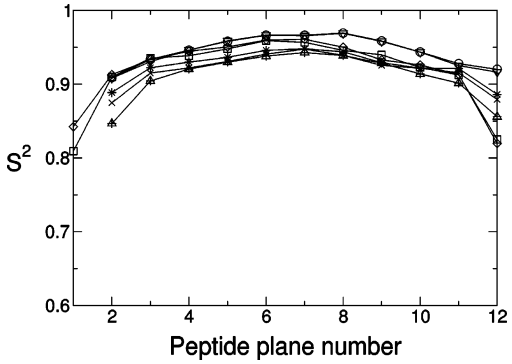

Figure 4. The iRED S^2 value plotted as a function of peptide plane number for the double trimmed ensemble of 5848 structures that meet the hydrogen bonding and end-to-end distance constraints. The symbols are the same as those used in Figure 1.

proteins can show small, additional features arising from tertiary interactions that are not captured by the current trimming criteria. The trimming procedure is also consistent with the molecular dynamics (MD) simulation of ubiquitin in that it accurately reflects the effects of reorientational motion on the different bond vector types. Average order parameters for each bond vector type were calculated from both the MD simulation of ubiquitin and the double trimmed (DT) model. The average values from the MD simulation exhibit the following trend:

$$\langle S^2_{\mathrm{N-C\alpha}}\rangle_{\mathrm{MD}} > \langle S^2_{\mathrm{C\alpha-C'}}\rangle_{\mathrm{MD}} > \langle S^2_{\mathrm{N-C'}}\rangle_{\mathrm{MD}} > \langle S^2_{\mathrm{C\alpha-C\beta}}\rangle_{\mathrm{MD}} > \langle S^2_{\mathrm{C\alpha-H\alpha}}\rangle_{\mathrm{MD}} > \langle S^2_{\mathrm{C'-O}}\rangle_{\mathrm{MD}} > \langle S^2_{\mathrm{N-H}}\rangle_{\mathrm{MD}} > \langle S^2_{\mathrm{(N-H)\times(N-C\alpha)}}\rangle_{\mathrm{MD}}$$

which is largely consistent with that observed from the double trimmed ensemble

$$\langle S^2_{\mathrm{N-C\alpha}}\rangle_{\mathrm{DT}} > \langle S^2_{\mathrm{C\alpha-C'}}\rangle_{\mathrm{DT}} > \langle S^2_{\mathrm{(N-H)\times(N-C\alpha)}}\rangle_{\mathrm{DT}} > \langle S^2_{\mathrm{C\alpha-C\beta}}\rangle_{\mathrm{DT}} > \langle S^2_{\mathrm{N-C'}}\rangle_{\mathrm{DT}} > \langle S^2_{\mathrm{C\alpha-H\alpha}}\rangle_{\mathrm{DT}} > \langle S^2_{\mathrm{N-H}}\rangle_{\mathrm{DT}} > \langle S^2_{\mathrm{C'-O}}\rangle_{\mathrm{DT}}$$

The main difference between the two ensembles is the average order parameter for the cross-product vectors (N−H) × (N−C$^\alpha$). A factor that contributes to the low $\langle S^2_{\mathrm{(N-H)\times(N-C\alpha)}}\rangle_{\mathrm{MD}}$ value is the fact that H^N atom experiences an out-of-plane motion that is absent in the double trimmed ensemble. An alternative cross-product vector (C′−O) × (N−C$^\alpha$) is less susceptible to this effect. For example, the average order parameter over the central region of the ubiquitin α-helix (peptide planes 27−31) for the (N−H) × (N−C$^\alpha$) vector is 0.90, whereas the corresponding value for the (C′−O) × (N−C$^\alpha$) vectors is 0.93.

The trimming procedure demonstrates that structures can exhibit relatively large amplitude fluctuations, yet still retain specific geometric features (H-bonds, end-to-end distance) provided that the dihedral angles are suitably correlated. To explore such correlations, Figure 5 plots the correlation coefficients between ϕ_7/ϕ_{7+k} (a) and ϕ_7/ψ_{7+k} (b) as a function of the relative amino-acid position k for the double trimmed 5848 conformer ensemble. Figure 5 also shows the corresponding correlations that are observed in the molecular dynamics (MD) simulation of ubiquitin. Dihedral angle ϕ_{29} was selected for this comparison as it is located in the middle of ubiquitin's central helix. Neither the double trimmed ensemble, nor the MD-derived ensemble exhibits significant ϕ/ϕ correlations. Both ensembles, however, exhibit a crankshaft motion between ϕ_i and ψ_{i-1}. In addition to this crankshaft motion, there is a relatively large ϕ_i/ψ_i anticorrelation. More distant correlations are weak and are damped out relatively quickly. The ϕ_i/ψ_i correlation has been

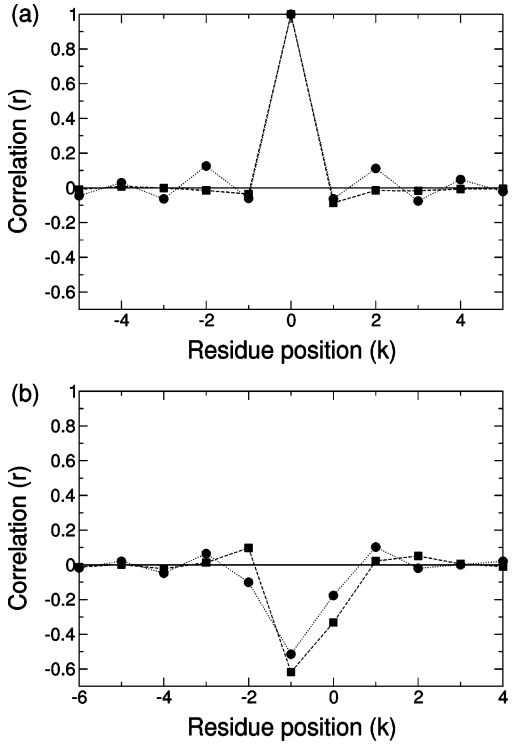

Figure 5. The correlation coefficients between ϕ_7/ϕ_{7+k} (a) and ϕ_7/ψ_{7+k} (b) plotted as a function of k for the double trimmed ensemble (circles). The correlation coefficients between ϕ_{29}/ϕ_{29+k} (a) and ϕ_{29}/ψ_{29+k} (b) plotted as a function of k from an MD simulation of ubiquitin (squares).

noted previously,[56,57] although it is not as widely recognized as the correlation between ϕ_i and ψ_{i-1}. The MD results indicate that the ϕ_i/ψ_i anticorrelation is most prevalent in α-helices and far less pronounced in β-sheets. This trend is not unique to ubiquitin, but has also been observed in MD simulations of other proteins that we have investigated, including MDM2 and RBD1 (unpublished). These results indicate that, while the β-sheet structure is maintained largely by ϕ_i/ψ_{i-1} crankshaft motions, the α-helical structure exhibits an additional ϕ_i/ψ_i correlation.

The absence of significant long-range correlations suggests that it might be possible to predict the S^2 values using a small set of local variances and covariances.[47] In Figure 6a, the variances $\mu_{\phi_i\phi_i}$ and $\mu_{\psi_i\psi_i}$ and covariances $\mu_{\phi_i\psi_{i-1}}$, $\mu_{\phi_i\psi_i}$, and $\mu_{\psi_i\psi_{i-1}}$ are plotted as a function of residue number. The variances and covariances that are observed in the N- and C-terminal residues are very close to the original values in the untrimmed ensemble. The central residues, however, are perturbed from their original values. The average variances and covariances across the five central residues (5−9) are $\langle\mu_{\phi_i\phi_i}\rangle = 65$ (deg^2), $\langle\mu_{\psi_i\psi_i}\rangle = 81$ (deg^2), $\langle\mu_{\phi_i\psi_{i-1}}\rangle = -35$ (deg^2), $\langle\mu_{\phi_i\psi_i}\rangle = -11$ (deg^2), and $\langle\mu_{\psi_i\psi_{i-1}}\rangle = -14$ (deg^2). As a comparison, the MD simulation of ubiquitin gives values of $\langle\mu_{\phi_i\phi_i}\rangle = 77$ (deg^2), $\langle\mu_{\psi_i\psi_i}\rangle = 70$ (deg^2), $\langle\mu_{\phi_i\psi_{i-1}}\rangle = -43$ (deg^2), $\langle\mu_{\phi_i\psi_i}\rangle = -28$ (deg^2), and $\langle\mu_{\psi_i\psi_{i-1}}\rangle = 0.3$ (deg^2). The double trimmed and MD-derived ensembles again provide a consistent overall trend, particularly in the largest variances and covariances $\mu_{\phi_i\phi_i}$, $\mu_{\psi_i\psi_i}$, and $\mu_{\phi_i\psi_{i-1}}$. Using this set of variances and covariances, the order parameter is expressed as

$$S_i^2 = 1 - c_{\psi_{i-1}\psi_{i-1}}\mu_{\psi_{i-1}\psi_{i-1}} - c_{\phi_i\phi_i}\mu_{\phi_i\phi_i} - c_{\psi_{i-1}\phi_i}\mu_{\psi_{i-1}\phi_i} \quad (6)$$

where the 3 coefficients $c_{\psi_{i-1}\psi_{i-1}}$, $c_{\phi_i\phi_i}$, and $c_{\psi_{i-1}\phi_i}$ are determined by a linear least-squares optimization using singular value decomposition with the iRED S_i^2 values and the μ values given

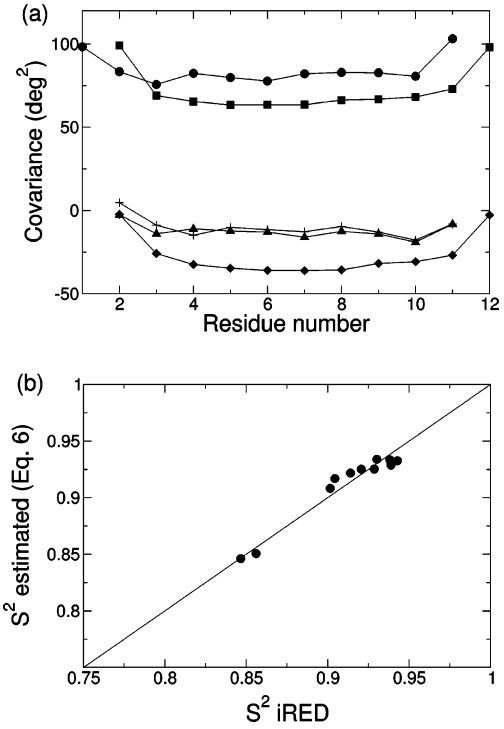

Figure 6. (a) The variances $\mu_{\phi_i\phi_i}$ (squares) and $\mu_{\psi_i\psi_i}$ (circles) and covariances $\mu_{\phi_i\psi_{i-1}}$ (diamonds), $\mu_{\phi_i\psi_i}$ (+), and $\mu_{\psi_i\psi_{i-1}}$ (up triangle) plotted as a function of residue number for the double trimmed ensemble. (b) A correlation plot between the S^2 value estimated from eq 6 and the iRED S^2 value for the N−H vector.

in Figure 6a. A correlation plot of the iRED S_i^2 values and the values obtained from the first-order expansion in eq 6 is shown in Figure 6b for the N−H bond vectors using $c_{\psi_{i-1}\psi_{i-1}} = -4.6 \times 10^{-4}$,$c_{\phi_i\phi_i} = 2.0 \times 10^{-3}$, $c_{\psi_{i-1}\phi_i} = 7.1 \times 10^{-4}$. Although the differences between residues are relatively subtle in Figure 6a, the variances and covariances represented therein provide an accurate estimate of the order parameter. The rmsd between the two sets is 0.007, while the linear correlation coefficient is 0.97. The inclusion of additional terms improves the fit only minimally. For example, the addition of the $c_{\psi_i\psi_i}\mu_{\psi_i\psi_i}$ and $c_{\phi_i\psi_i}\mu_{\phi_i\psi_i}$ terms decreases the rmsd to 0.005 and increases the correlation coefficient to 0.99. Equation 6 captures the order parameters for the MD-derived ensemble of ubiquitin well (rmsd of 0.02 and linear correlation coefficient of 0.93), after the 3 coefficients of eq 6 have been suitably adjusted.

3D GAF Order Parameters. Peptide planes of the ensembles studied here generally do not reorient "symmetrically" as is manifested in the order parameter differences within the individual peptide planes (Figures 2 and 4). Additional insight into the factors governing these differences is provided by a 3D GAF analysis. In Figure 7a, σ_α, σ_β, and σ_γ (collectively referred to as σ_{3DGAF}) are plotted as a function of peptide plane number for the double trimmed ensemble. The trend is similar to that observed among the variances and covariances in Figure 6a in that the fluctuation amplitudes increase near the termini. Note, however, that a correspondence between Figures 6a and 7a is not necessary. As an example, consider Figure 7b in which the σ_{3DGAF} profile is provided for the ensemble in which the ϕ_i/ψ_{i-1} correlation coefficient is explicitly set to −0.75. Recall that this ensemble yields a similar S^2 profile as the double trimmed ensemble. A comparison of Figures 7a and 7b indicates that the σ_{3DGAF} profiles for the two ensembles share similar features. Despite these similarities, the two ensembles exhibit very different fluctuations in their internal (dihedral angle) coordinates. The trimmed ensemble exhibits variations across

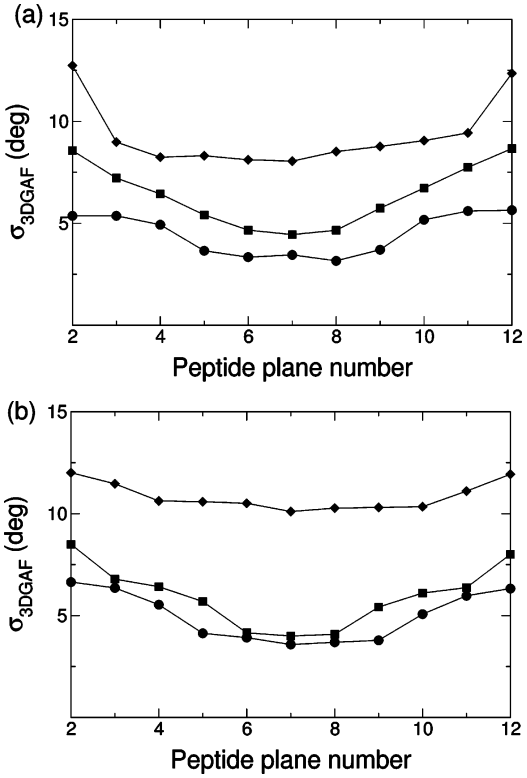

Figure 7. The 3D GAF fluctuation amplitudes σ_α (circles), σ_β (squares), and σ_γ (diamonds) plotted as a function of peptide plane number for the double trimmed (a) and $r = -0.75$ (b) ensembles.

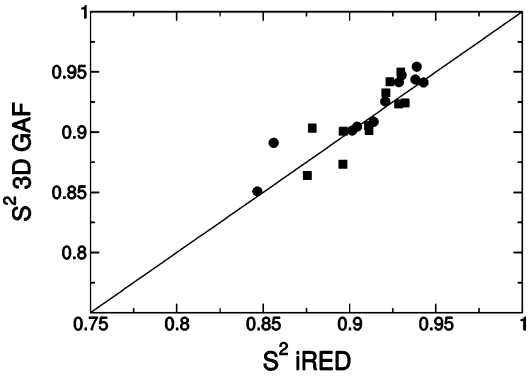

Figure 8. A correlation plot between the 3D GAF and iRED S^2 values for the N−H vector in the double trimmed (circles) and $r = -0.75$ (squares) ensembles.

the sequence that are missing in the $r = -0.75$ ensemble. In the $r = -0.75$ ensemble, the variances and covariances in ϕ/ψ are uniform ($\mu_{\phi_i\phi_i} = \mu_{\psi_i\psi_i} = 100$ deg^2, $\mu_{\phi_i\psi_{i-1}} = -75$deg^2). This observation implies that the S^2 variations in the $r = -0.75$ ensemble are due to longer range motions that propagate over multiple residues. The σ_{3DGAF} values are sensitive not only to these long-range motions, but also to short-range motions that modulate the local internal coordinates; thus 3D GAF motions include both of these contributions. They represent a cumulative effect in which a large number of motional modes are projected onto the individual peptide planes. As result, ensembles with rather different ϕ/ψ fluctuations can yield similar σ_{3DGAF} (and S^2) profiles.

Order parameters are calculated in the 3D GAF analysis from the fluctuation amplitudes and the mean bond vector orientation in the 3D GAF reference frame. A correlation plot of the iRED and 3D GAF S^2 values is provided in Figure 8 for the N−H bond vectors. The double trimmed and $r = -0.75$ ensembles

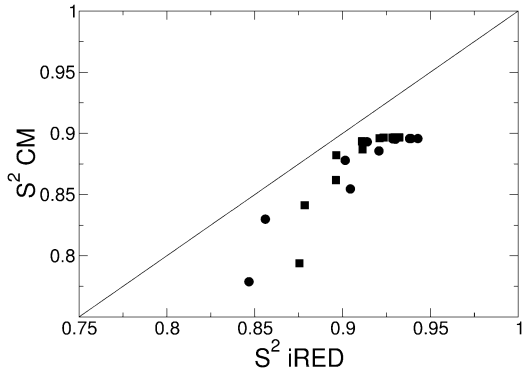

Figure 9. A correlation plot between the contact model and iRED S^2 values for the N−H vector in the double trimmed (circles) and $r = -0.75$ (squares) ensembles.

are represented as circles and squares, respectively. The correlation between the iRED and 3D GAF methods is high in both cases. The double trimmed ensemble gives an rmsd of 0.014 and a linear correlation coefficient of 0.94, while the $r = -0.75$ ensemble gives an rmsd of 0.015 and a correlation coefficient of 0.82. The largest discrepancies between the two methods are observed among the two to three peptide planes toward the N− or the C−terminus.

Local Contact Model Calculations. A local contact model can account for many features of backbone order parameters in globular proteins.[29] The contact model is generally applied either to a high-resolution X-ray structure or to a small ensemble of NMR structures. In the present context, the local contact model is applied to all members of the helical ensemble and the resulting S^2 values are averaged over the ensemble. A comparison between the N−H order parameters by the local contact model and the ones by the iRED method is given in Figure 9. The results for the double trimmed and $r = -0.75$ ensembles are represented by circles and squares, respectively. The double trimmed ensemble gives a correlation coefficient of 0.94 and the $r = -0.75$ ensemble a value of 0.88. The contact model exhibits a plateau at $S^2 = 0.9$, which reflects the experimental S^2 values that are typically observed in a rigid protein segment. The helical peptide system exhibits a slightly higher baseline order parameter, indicating that it is not entirely representative of globular proteins for which the local contact model was parametrized. As noted above, it lacks zero-point vibrational contributions that affect the order parameter baseline.[19,20]

Conclusions

An S^2 order parameter represents quantitative information about the reorientational fluctuation of an interaction vector at an individual site. Because of its general mathematical nature, a given S^2 value can be consistent with a large number of different possible motional models. This ambiguity can be overcome by placing S^2 into context: the present work demonstrates how an extensive set of order parameters belonging to a helical segment reflects upon structural and dynamic properties and constraints, such as end-to-end distance constraints, hydrogen bonds, local contacts, locally anisotropic dynamics, and local backbone φ,ψ dihedral angle motional amplitudes and their mutual correlations. This information provides the basis to address the inverse problem, which is the extraction of this kind of structural and dynamic information from experimentally accessible order parameters. The minimalistic assumptions made for the generation of the helical ensembles and the fact that many of the structural dynamic features observed mirror those found in all-atom MD simulations in explicit solvent suggest that these patterns are of a general nature and therefore rather broadly applicable to helices in proteins.

Order parameters have been calculated here for ensembles of a helical model peptide by five complementary methods: (1) the standard eq 1, (2) iRED, (3) a first-order expansion in local variances and covariances, (4) the 3D GAF method, and (5) the local contact model. Despite their differences, all five methods provide consistent results for the scenarios considered here. This observation is noteworthy given the fact that the order parameter variations are relatively small in several of the ensembles. Each method offers complementary information. (1) The standard definition provides a useful reference. (2) The iRED method allows one to assess separability between internal and overall rotation, which is a basic assumption of the model-free approach. (3) The first-order expansion specifically accounts for the contributions from local dihedral angle fluctuations. (4) The 3D GAF model provides a full description of anisotropic peptide plane motion, and (5) the local contact model demonstrates an intrinsic link between protein structure and dynamics. By applying all five of these methods to a simple model system whose dynamics is governed by minimal assumptions, one obtains a comprehensive view of the motional processes that give rise to order parameter variations. For this purpose, it is instructive to note that method 3 is unique in that it is insensitive to contributions from long-range motion. The fact that method 3 is successful when applied to the double trimmed ensemble indicates that long-range motions do not strongly contribute to the order parameters in that case. The structural constraints that are present in the double trimmed ensemble introduce local variations in the ϕ/ψ fluctuations. These effects, which are local in origin, are reflected in the order parameter variations across the sequence. This scenario corresponds to what is commonly observed in the core of globular proteins. The situation is quite different, however, for the ensembles in which the ϕ_i/ψ_{i-1} correlation is explicitly defined. In these cases, method 3 fails to describe changes of S^2 along the helix due to the uniformity of the local ϕ/ψ fluctuations. Variations in the order parameter, therefore, by necessity reflect contributions from long-range motion. This effect is included in the other four methods, which project the contribution from long-range motion onto the local environment. For example, the contact model indicates that the number of local contacts decreases near the termini for the $r = -0.75$ ensemble. Likewise, the 3D GAF model demonstrates that the σ_{3DGAF} values increase near the termini. Although the origin of the effect is nonlocal, it is seen locally via S^2. These factors give rise to the fraying effect that is often observed in the order parameter profiles of globular proteins toward the termini of the polypeptide chain. The ability to discriminate between these various scenarios is one of the benefits of the approach presented here. This case study demonstrates that a self-consistent view of reorientational motion emerges when multiple complementary methods are applied to the same system. Generalization of the properties for an α-helix described here to other structural motifs will further enhance the power of NMR order parameter analysis of proteins.

Acknowledgment. E.J. is the recipient of a postdoctoral fellowship from the American Heart Association, Greater Southeast Affiliate. S.A.S. is the recipient of an NIH postdoctoral fellowship. This work was supported by the NSF (Grant 0621482).

References and Notes

(1) Palmer, A. G. *Chem. Rev.* **2004**, *104*, 3623.
(2) Brüschweiler, R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 175.
(3) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224.
(4) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546.
(5) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559.
(6) Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989**, *28*, 8972.
(7) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098.
(8) Tolman, J. R. *J. Am. Chem. Soc.* **2002**, *124*, 12020.
(9) Jarymowycz, V. A.; Stone, M. J. *Chem. Rev.* **2006**, *106*, 1624.
(10) Case, D. A. *Acc. Chem. Res.* **2002**, *35*, 325.
(11) Zhuravleva, A.; Korzhnev, D. M.; Nolde, S. B.; Kay, L. E.; Arseniev, A. S.; Billeter, M.; Orekhov, V. Y. *J. Mol. Biol.* **2007**, *367*, 1079.
(12) Wang, W.; Nymeyer, H.; Zhou, H.-X.; Berg, B.; Brüschweiler, R. *J. Comput. Chem.*
(13) Chen, J.; Brooks, C. L.; Wright, P. E. *J. Biomol. NMR* **2004**, *29*, 243.
(14) Markwick, P. R.; Bouvignies, G.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 4724.
(15) Nederveen, A. J.; Bonvin, A. M. J. J. *J. Chem. Theory Comput.* **2005**, *1*, 363.
(16) Palmer, A. G.; Case, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 9059.
(17) Levy, R. M.; Karplus, M.; Wolynes, P. G. *J. Am. Chem. Soc.* **1981**, *103*, 5998.
(18) Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961.
(19) Case, D. A. *J. Biomol. NMR* **1999**, *15*, 95.
(20) Brüschweiler, R. *J. Am. Chem. Soc.* **1992**, *114*, 5341.
(21) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325.
(22) Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules* **1984**, *17*, 1370.
(23) Brüschweiler, R.; Case, D. A. *Phys. Rev. Lett.* **1994**, *72*, 940.
(24) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 7305.
(25) Pfeiffer, S.; Fushman, D.; Cowburn, D. *J. Am. Chem. Soc.* **2001**, *123*, 3021.
(26) Haliloglu, T.; Bahar, I. *Proteins* **1999**, *37*, 654.
(27) Ming, D.; Brüschweiler, R. *Biophys. J.* **2006**, *90*, 3382.
(28) Abergel, D.; Bodenhausen, G. *J. Chem. Phys.* **2005**, *123*, 204901.
(29) Zhang, F.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 12654.
(30) Ming, D.; Brüschweiler, R. *J. Biomol. NMR* **2004**, *29*, 363.
(31) Bremi, T.; Brüschweiler, R. *J. Am. Chem. Soc.* **1997**, *119*, 6672.
(32) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870.
(33) Akke, M.; Brüschweiler, R.; Palmer, A. G. *J. Am. Chem. Soc.* **1993**, *115*, 9832.
(34) Prompers, J. J.; Brüschweiler, R. *J. Phys. Chem. B* **2000**, *104*, 11416.
(35) Yang, D.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369.
(36) Wand, A. J. *Nat. Struct. Biol.* **2001**, *8*, 926.
(37) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. *Nature* **2007**, *448*, 325.
(38) Peng, J. W.; Wagner, G. *Biochemistry* **1995**, *34*, 16733.
(39) Peng, J. W.; Wagner, G. *Biochemistry* **1992**, *31*, 8571.
(40) Peng, J. W.; Wagner, G. *J. Magn. Reson.* **1992**, *98*, 308.
(41) Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6439.
(42) Skrynnikov, N. R.; Millet, O.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6449.
(43) Meirovitch, E.; Shapiro, Y. E.; Polimeno, A.; Freed, J. H. *J. Phys. Chem. A* **2006**, *110*, 8366.
(44) Meirovitch, E.; Shapiro, Y. E.; Polimeno, A.; Freed, J. H. *J. Phys. Chem. B* **2007**, *111*, 12865.
(45) Vugmeyster, L.; Raleigh, D. P.; Palmer, A. G., III; Vugmeister, B. E. *J. Am. Chem. Soc.* **2003**, *125*, 8400.
(46) Goodman, J. L.; Pagel, M. D.; Stone, M. J. *J. Mol. Biol.* **2000**, *295*, 963.
(47) Brüschweiler, R. *J. Chem. Phys.* **1995**, *102*, 3396.
(48) MacKerell, A. D.; Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; III;, Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(49) Bremi, T.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1997**, *119*, 4272.
(50) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522.
(51) Showalter, S. A.; Baker, N. A.; Tang, C.; Hall, K. B. *J. Biomol. NMR* **2005**, *32*, 179.
(52) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Brüschweiler, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885.
(53) Ferrage, F.; Pelupessy, P.; Cowburn, D.; Bodenhausen, G. *J. Am. Chem. Soc.* **2006**, *128*, 11072.
(54) Henry, E. R.; Szabo, A. *J. Chem. Phys.* **1985**, *82*, 4753.
(55) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.
(56) Levy, R. M.; Karplus, M. *Biopolymers* **1979**, *18*, 2465.
(57) Go, M.; Go, N. *Biopolymers* **1976**, *15*, 1119.
(58) Chandrasekhar, I.; Clore, G. M.; Szabo, A.; Gronenborn, A. M.; Brooks, B. R. *J. Mol. Biol.* **1992**, *226*, 239.
(59) Fadel, A. R.; Jin, D. Q.; Montelione, G. T.; Levy, R. M. *J. Biomol. NMR* **1995**, *6*, 221.
(60) Wang, T. Z.; Cai, S.; Zuiderweg, E. R. P. *J. Am. Chem. Soc.* **2003**, *125*, 8639.