Dynamical characterization of residual and non-native structures in a partially folded protein by ^{15}N NMR relaxation using a model based on a distribution of correlation times

FRANÇOISE OCHSENBEIN,^{1,3} JEAN-MICHEL NEUMANN,² ERIC GUITTET,¹ AND CARINE VAN HEIJENOORT¹

¹CNRS, Institut de Chimie des Substances Naturelles, F-91190 Gif sur Yvette, France

²Département de Biologie Cellulaire et Moléculaire, Section de Biophysique des Protéines et des Membranes, F-91191 Gif sur Yvette Cedex, France

(RECEIVED October 1, 2001; FINAL REVISION December 8, 2001; ACCEPTED December 8, 2001)

Abstract

A spectral density model based on a truncated lorentzian distribution of correlation times is used to analyze the nanosecond time-scale dynamics of the partially unfolded domain 2 of annexin I from its ^{15}N NMR relaxation parameters measured at three magnetic field strengths. The use of a distribution of correlation times enables the characterization of the dynamical features of the NH bonds of the protein in terms of heterogeneity of dynamical states in the nanosecond range. The variation along the sequence of the two dynamical parameters introduced, namely the center and the width of the distribution, points out the different types of residual secondary structures present in the D2 domain. Moreover, it allows a physically sensible interpretation of the dynamical behavior of the different residual helices and of the non-native structures. Also, a striking correspondence is found between the parameters obtained using an extended Lipari and Szabo model and the parameters obtained using the distribution of correlation times. This result led us to propose a specific interpretation of the model-free order parameter for internal motions in the nanosecond range in the case of unfolded states.

Keywords: Distribution of correlation times; dynamics; models; proteins; ^{15}N relaxation; unfolded states

In addition to their long-known importance in folding processes, unstructured proteins increasingly appear to play important roles in biological processes (for review, see Dyson and Wright 2001). This reinforces the interest for a detailed characterization of unfolded or partially folded states of proteins. In contrast to fully folded proteins that can generally be described as a well-defined structure with

limited fluctuations, unfolded or partially folded proteins are highly dynamical and consist of a wide range of interconverting conformers. This is nicely illustrated by a recent study of an unfolded state comprising residual helices (Korzhnev et al. 1999a, 1999b; Orekhov et al. 1999). The resolution of kinetic equations relative to a helical peptide led to a nearly continuous distribution of transition times from 0.1 to 10 ns. These frequencies are associated with interconversion between various helical states, such as the elongation or shortening of the helix segment by several residues or interconversion between α , π , and 3/10 helical conformations. Within this continuum of frequencies it seems difficult to distinguish separated frequency ranges as in the case of folded proteins. Conversely, the inherent high flexibility of unfolded states makes the knowledge of their dynamical

Reprint requests to: Françoise Ochsenbein, CNRS, Institut de Chimie des Substances Naturelles, 1 avenue de la terrasse, F-91190, Gif sur Yvette, France; e-mail: carine@icsn.cnrs-gif.fr; fax: 33-(0)1-69823784.

³Present address: Département de Biologie Cellulaire et Moléculaire, Section de Biophysique des Protéines et des Membranes, CEA de Saclay and URA CNRS 2096, F-91191 Gif sur Yvette Cedex, France.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.4000102>.

behavior a crucial point for their characterization. NMR spectroscopy remains the most efficient technique to obtain detailed information on the residual structure and on the dynamics of unfolded states (Alexandrescu and Shortle 1994; Frank et al. 1995; Buck et al. 1996; Brutscher et al. 1997; Farrow et al. 1997; Schwalbe et al. 1997; Dyson and Wright 1998; Eliezer et al. 1998, 2000; Meekhof and Freund 1999; Ochsenbein et al. 2001). However, it appears from these studies that, in the case of unfolded proteins, it is difficult to specify and/or interpret the parameters derived from the commonly used model-free approach. The extraction of motion parameters from models based on discrete correlation times, such as the model-free approaches, presents two major difficulties. The first major problem lies in the consideration of a global correlation time, representing the slowest time leading to the complete loss of correlation of motions, which is common to every residue of the protein. This idea of a unique overall correlation time for the whole molecule is questionable in the case of unfolded states. An attempt to alleviate this problem is to use a so-called "local model-free" approach, which consists in fitting one "global" correlation time per residue (Alexandrescu and Shortle 1994). The resulting correlation times are then associated with the reorientation of segments of the protein that have independent dynamical behaviors. Although this approach led to some acceptable results, especially in the case of proteins containing fully folded segments and unfolded tails (Brutscher et al. 1997; van Heijenoort et al. 1998), the independence of the dynamical behavior of each segment and thus the interpretation of the motional parameters obtained remain questionable. The second major problem of the model-free approach is the analysis of internal motions, because they are considered as a whole, without *a priori* about the type of motions they represent. It is usually accepted for folded proteins that an increase of the order parameter corresponds to a decrease of the amplitude of motions around a given conformation (wobbling in a cone) or between two or more given conformations (jump model) (Lipari and Szabo 1982). However, this physical interpretation may be inappropriate when considering an unfolded protein with no defined structure. Moreover, if motions over a continuous range of frequencies are taking place, the question of the physical meaning of an effective internal correlation time remains open.

A simple way to deal with the notion of motions over a continuum of frequency in the nanosecond range is to introduce a continuous distribution of correlation times in the autocorrelation function. A number of such empirical functions were already proposed in the context of synthetic polymers that share many common features with unfolded states of proteins (for review, see Heatley 1979, 1986). One of these functions, the Cole–Cole distribution was recently applied by Buevich et al. on the unfolded pro-peptide of subtilisin PPS (Buevich and Baum 1999; Buevich et al. 2001).

A major limitation of these functions is that they were designed to fit only experimental longitudinal ^{13}C relaxation rate constants and heteronuclear nOes, which are related to spectral densities in the megahertz range. They were not optimized to fit transverse relaxation rate constants, which are related to the spectral density in the kilohertz range, and which are not relevant to the studies of fast internal dynamics of very large polymers. The derived spectral density function then usually diverges when ω tends toward zero. The direct use of such distribution functions is problematic in the case of small proteins for which transverse relaxation rate constants contain a high contribution of nanosecond time-scale motions and are sensitive to fast internal motions. We propose here the use of a lorentzian distribution of correlation times, characterized by its center τ_0 and its width Δ , and extending from zero to an upper value τ_{max} . The physical meaning of τ_{max} , which is introduced to ensure convergence, is discussed.

We applied this approach to the analysis of the ^{15}N relaxation data obtained at three magnetic fields, 500, 600, and 800 MHz, on the partially folded D2 domain of annexin I. This domain is unable to adopt a tertiary fold alone in solution. Conversely, it is fully folded in the entire annexin I, in which it exhibits five helix segments, named A to E, organized in the superhelix motif characteristic of annexin domains (Huber et al. 1992). A structural analysis of this domain was previously performed that showed that the residual structure of the isolated domain 2 of annexin I comprises both native and non-native secondary structure elements (Cordier-Ochsenbein et al. 1996, 1998). The native elements consist of four helix segments corresponding to the B and E native helices as well as a large proportion of the A helix and a small helix population corresponding to the second part of the C helix. On both sides of the residual C helix, the conformation of the polypeptide chain consists in a set of interconverting turn-like structures that prevents the formation of the remaining elements of the native C helix. Finally, the N and C-terminal region and the segment spanning the native D helix remain random coil. These residual structures are illustrated in Figure 1.

We address the question of the dynamical behavior of these different types of residual structures. It should be noted that their coexistence in the same protein makes the isolated domain D2 of annexin I an attractive model for the investigation of the dynamics of partially folded states. A "model-free" analysis of the relaxation data previously performed on this protein led to ambiguous results with a local and extended Lipari and Szabo model being both statistically satisfactory. This emphasized the difficulty to interpret the discrete correlation times and the order parameters obtained from the model-free approach (Ochsenbein et al. 2001). We show here that the use of the lorentzian distribution of correlation times leads to a relevant representation of the dynamical features of the D2 domain. It allows us to

Fig. 1. (A) Ribbon representation of the structure of domain 2 of annexin I extracted from the crystal structure of the entire protein (Weng et al. 1993) (*top view*). (B) Schematic representation of the structural features of isolated domain 2 of annexin I in solution from NMR data (*side view*). Residual helices are represented as ribbons, regions showing non-native structures are represented in black, unfolded regions are shown in light gray. These figures were drawn using the Molscript program (Kraulis 1991).

discriminate the various types of residual structures of the isolated D2 domain. Values of the center of the distribution of correlation times enable to distinguish the intrinsically stable helices A, B, and C from helix E. In addition, high values of the width of the distribution highlight the heterogeneous dynamical behavior of the interconverting turn-like structures. These results also lead to favor one of the two model-free approaches that fit experimental relaxation data, and bring a physical interpretation, relevant for unfolded states, of the generalized order parameter in term of state heterogeneity.

Results and Discussion

Calculation of the spectral density function for a lorentzian distribution of correlation times

The general expression of the spectral density function derived from the use of a distribution of correlation times is given in the Materials and Methods section (equations 1a–c). When considering a distribution of correlation times, the spectral density is a continuous function of both the frequency ω and the correlation time τ , $J(\omega, \tau)$ (equation 1c). The general expression of the spectral density function $J(\omega)$ is then obtained from the integration along the correlation time dimension of $J(\omega, \tau)$ weighted by the function of distribution of correlation times (equation 1a) (Heatley 1979). In the special case of symmetrical functions of distribution, such as the lorentzian used here (equation 1b), the resulting spectral density $J(\omega)$ becomes a function of two time pa-

rameters: the center of the distribution τ_0 , and the width of the distribution Δ .

One difficulty inherent to the use of continuous distributions of correlation times $F(\tau)$ arises from the fact that they extend toward infinity. The integration of $F(\tau)J(\omega, \tau)$ up to infinity then generates large values of the spectral density at low frequency, due to a contribution of the long correlation times that becomes significant. Especially, the integral diverges towards infinity for $J(\omega = 0)$. These slow motions do not contribute to the NMR experimental relaxation data that contain only the contribution of motions in the ps–ns range, and eventually a contribution due to conformational exchange in the μ s–ms range. This last one can be evaluated from the field dependence of $J(0)$ (Phan et al. 1996), cross-correlation rates measurements (Kroenke et al. 1998) or from off-resonance experiments (Mulder et al. 1998; Desvaux and Berthault 1999) and subtracted from the experimental transverse relaxation times. To avoid the divergent behavior of the spectral density function, it is necessary to truncate the distribution function. We thus introduced the parameter $\tau_{\max}$, which determines the upper limit of the distribution of correlation times (equation 2b). This parameter corresponds to the slower correlation time leading to the complete loss of correlation of motions. It should be noted that this slower correlation time is not well determined for unfolded states of proteins. It is likely to be much higher than the overall correlation time of a folded protein of the same molecular weight. In the fitting procedure of experimental data, several values of $\tau_{\max}$ going from 20 to 500 ns were tested. Small values of $\tau_{\max}$ led to erroneous values of τ_0 for some residues (nearly zero with very high uncertainties), whereas too high values of $\tau_{\max}$ increased the value of χ^2_{exp} . Also, changes in the value of $\tau_{\max}$ did not perturb the relative evolution of the parameters along the sequence of the protein. We thus considered that a value of 100 ns was a reasonable upper limit of the distribution of correlation times. This value would correspond to a maximal gyration radius of 5.2 nm for the protein at 308 K (using a classical Stoke-Einstein relation

$$R = \sqrt[3]{\tau_c \frac{3kT}{4\pi\eta}},$$

T being the temperature in Kelvin, η the viscosity, and k the Boltzman constant), about 2.5 times more than the gyration radius of a folded protein of this size. This maximum gyration radius is comparable to the largest dimension measured for the molten globules of apomyoglobin and alpha-lactalbumin by X-ray scattering scaled by the molecular weight ratio (Kataoka et al. 1995, 1997).

It should be noted that another distribution of correlation times, the Cole–Cole distribution, developed in the context of the analysis of longitudinal ^{13}C relaxation times of polymers, was recently applied to the unfolded protein PSS

(Buevich and Baum 1999; Buevich et al. 2001). The same problem of divergence occurs when using this distribution. In this case, the problem lies in the choice of the lower limit of the spectral density function, because $R_N(N_{x,y})$ values back-calculated during the fitting procedure cannot be obtained from $J(0)$, which is infinite. Ignored in the first report by the authors (Buevich and Baum 1999), this problem was discussed in some details in the second report (Buevich et al. 2001) in which the authors replaced $J(0)$ by the value of the spectral density computed at a frequency equal to one $\text{rad}\cdot\text{s}^{-1}$.

Determination of the dynamical parameters

The experimental data, consisting of the nine relaxation parameters, $R_N(N_z)$, $R_N(N_{x,y})$ and heteronuclear (^1H - ^{15}N) nOes obtained on the partially folded domain 2 of annexin I at three magnetic fields (500, 600, and 800 MHz), were measured as previously described (Ochsenbein et al. 2001). The contribution of slow conformational exchange processes in the μs – ms range was evaluated using the method described by Phan et al. (1996), and was removed from the experimental transversal relaxation rate constants (Ochsenbein et al. 2001).

The simple model considering only a lorentzian distribution of correlation times (equation 3) was not able to reproduce correctly experimental data, that is, the sum of the squared residual errors for the relaxation parameters of each N—H, χ^2_{exp} , was more than the critical value $\chi^2_{95\%}$ obtained from the Monte Carlo simulations. It was then extended to take into account the possible presence of fast motions in the picosecond time scale (equation 4). A third parameter A_{LD} is thus introduced, which is related to the generalized order parameters S^2 or S^2_f of the model-free approach (Lipari and Szabo 1982; Clore et al. 1990). The need to consider libration motions had been previously reported in dynamical studies of polymers (Dejean de la Batie et al. 1988). The fit of the experimental data against this extended model using $J_{\text{LD2}}(\omega)$ (equation 4) fulfilled the criteria of goodness of fit for all residues, meaning that the sum of the squared residual errors for the relaxation parameters of each N—H, χ^2_{exp} , was lower than the critical value $\chi^2_{95\%}$ obtained from the Monte Carlo simulations. The resulting motional parameters are shown in Figure 2 (left panel).

Analysis of the dynamical parameters obtained on the partially folded D2 domain of annexin I

Values of A_{LD} associated with fast internal motions remain uniformly high along the whole backbone of the protein, with a mean value of 0.73 and a standard deviation of 0.03. This indicates that libration motions in the picosecond time-scale participate to around 25% of the total loss of correla-

tion of motions and do not depend on the degree of local structuration.

The center of the distribution τ_0 reflects the time scale of the most probable motions, and corresponds to fluctuations involving low-energy barriers. The width of the distribution Δ is related to the dynamical, and thus probably conformational, heterogeneity. From the observation of the evolution of τ_0 and Δ along the sequence, three types of behavior can be distinguished (Fig. 2, Table 1). The first type of behavior is characterized by the highest values of τ_0 ($1.4 \pm 0.2 \text{ ns}\cdot\text{rd}^{-1}$) and intermediate values of Δ . The residues following this first dynamical behavior are located in the residual A, B, and C helices. These values of τ_0 remain much shorter than those expected for the global reorientation of a domain of the size of the D2 domain. They are more compatible with the independent reorientation of segments of the size of a helix. This suggests that these helices are persistent and behave as independent units. The second type of behavior is characterized by the lowest values of both τ_0 and Δ . It involves residues in the unfolded regions and in the residual E helix. Homogeneous (low Δ) and faster (smaller τ_0) motions are indeed expected for residues in unstructured regions where local interactions are likely to be limited. The “unfolded type” dynamical behavior of the residues of E helix is more surprising. It indicates that the energy barriers are lower in this helix than in the residual A, B, and C helices. A number of experiments, including structural studies of fragments of the D2 domain, showed that helices A, B, and C have a significant intrinsic helix propensity, contrary to D and E helices (Cordier-Ochsenbein et al. 1998). In particular, according to chemical shift indexes, D and E helices are not formed in isolated short fragments. Because it was also shown that the segment corresponding to the E helix retains a large helical population in the isolated domain 2, it was proposed that this residual E helix is mainly stabilized by transient nonlocal hydrophobic interactions. Our results indicate that the inherent instability of the residual E helix is associated with rapid folding/unfolding rates, leading to correlation times close to those of an unfolded region. The last two regions, located at the extremities of the residual C helix, display a third dynamical behavior. In these two segments, values of τ_0 are comparable to those found in unfolded regions and in helix E, but values of the width of the distribution are the highest in the whole sequence ($0.7 \pm 0.2 \text{ ns}\cdot\text{rd}^{-1}$ compared to $0.4 \pm 0.2 \text{ ns}\cdot\text{rd}^{-1}$). These two regions were shown to adopt non-native structures consisting of interconverting turn-like structures stabilized by local specific interactions such as salt bridges and capping interactions (Cordier-Ochsenbein et al. 1996, 1998). The high values of Δ highlight a higher dynamical heterogeneity in these regions compared to the regions containing residual secondary structures, indicating a wider range of energy barriers for the interconversion between the various non-native turn-like structures. In these regions, the

Fig. 2. Dynamical parameters derived from ^{15}N relaxation measurements obtained on the D2 domain at pH 6 and 35°C as a function of the domain sequence. *Left panel:* lorentzian distribution model (equations 1–2), *middle panel:* extended model-free approach, *right panel:* local model-free approach. White boxes on top of the figure delineate the residual helices; gray boxes delineate the non-native turn-like structures. The value of the upper limit of the lorentzian distribution (τ_{max}) is equal to $100 \text{ ns}\cdot\text{rd}^{-1}$. Bars represent the standard deviation of simulated parameters. (A) A_{LD} is analogous to the order parameters S_f^2 and S^2 associated with fast internal motions in the extended Lipari and Szabo model and in the local model-free approach, respectively. (B) τ_0 corresponds to the center of the lorentzian distribution, τ_s corresponds to the effective internal correlation time for slow internal motions, τ_c^{local} is the slowest correlation time adjusted for each residue. (C) Δ corresponds to the width of the lorentzian distribution, S_s^2 corresponds to the generalized order parameter for internal slow motions, τ_i corresponds to the effective internal correlation time in the local model-free approach. (D) Quality criteria of the fit: the error function χ^2_{exp} (filled squares) results from the nonlinear fit of experimental data (see Materials and Methods). The fit was considered as satisfactory if χ^2_{exp} is ranged within the 95% confidence limit $\chi^2_{95\%}$ (continuous lines) obtained from 500 Monte Carlo simulations, that is, $\chi^2_{\text{exp}} < \chi^2_{95\%}$.

high dynamical heterogeneity appears to be associated with a large structural heterogeneity. In their study, Buevich et al. pointed out a relationship between the widths of the distri-

Table 1. Mean values of parameters for the various regions of the D2 domain of annexin I, obtained from the model using a lorentzian distribution of correlation times for each residue

	A, B and C residual helices	Non-native structures	Residual E helix	Unfolded residues
$\langle\tau_0\rangle$ ($\text{ns}\cdot\text{rd}^{-1}$)	1.4 ± 0.2	1.0 ± 0.2	0.9 ± 0.2	0.8 ± 0.3
$\langle\Delta\rangle$ ($\text{ns}\cdot\text{rd}^{-1}$)	0.5 ± 0.2	0.7 ± 0.2	0.4 ± 0.2	0.4 ± 0.2
$\langle A_{\text{LD}}\rangle$	0.75 ± 0.03	0.75 ± 0.03	0.76 ± 0.03	0.75 ± 0.04

bution in the nanosecond time scale and the presence of electrostatic interactions (Buevich et al. 2001). Especially, an analysis of the dependence of the width of the distribution on pH led them to propose that formation of ion pairs or hydrogen bonds upon a change of the ionic state of the protein yielded more correlated motions. This results in a larger distribution of states, and consequently, in larger widths of distribution of correlation times. The increase of Δ in the region of non-native structures around helix C in the D2 domain is in full agreement with this hypothesis. However, the authors propose to extend this hypothesis to all kind of inter-residue interactions (electrostatic, dipolar, and hydrophobic), and suggest a correlation between the width of the distribution and an adapted Kyte-Doolittle hy-

dropathy. It should be noted that the correlation found in the case of PPS is rather low (0.43). In our case, no significant correlation is found between the width of the distribution and the absolute values of the Kyte-Doolittle hydropathy (the correlation across the whole sequence is found equal to 0.29; data not shown). Moreover, in the case of the D2 domain, N—H bonds of residues located in stable residual secondary structures, which contain hydrogen bonds and are likely to undergo correlated motions, do not present a significant increase of the width of the distribution, but rather exhibit higher values of the center of the distribution. Our interpretation of the distribution parameters in terms of the presence and nature of residual secondary structures seems more adapted to the case of partially folded proteins.

At that point, it is important to compare these results with the more conventional model-free analysis. Figure 2 shows the motion parameters resulting from a model-free analysis of the D2 domain relaxation data. Two models containing three parameters were able to correctly reproduce the experimental data (Ochsenbein et al. 2001). The first model was an extended Lipari and Szabo model (Clore et al. 1990) with an optimized overall tumbling correlation time equal to $5 \text{ ns}\cdot\text{rd}^{-1}$ (Fig. 2, central panel). The second model was a simple Lipari and Szabo model (Lipari and Szabo 1982) in which the correlation time in the nanosecond range was adjusted on a per-residue basis (Fig. 2, right panel). Although both models were statistically equally relevant, they led to different motional parameters, emphasizing the difficulty to interpret the discrete correlation times and the order parameters obtained from the model-free approaches.

Figure 3 shows a comparison of the results obtained using the three approaches for three residues: Ala 11, Ile34, and Asp60. Ala 11 is located in the residual A helix and is representative of the residues located in the intrinsically stable helices. Ile 34 is located in one of the segments containing non-native structures. Asp 60 is located at the beginning of the residual E helix and is representative of residues of low helical propensity and of residues in unfolded segments. The figure highlights the localization of the discrete correlation times obtained from the two model-free approaches with respect to the lorentzian distribution of correlation times. The correlation times obtained from a simple model-free approach, consisting of one local correlation time τ_c per residue and an internal fast correlation time τ_i , appear to be located on each side of the optimized distribution of correlation times. On the contrary, the slow internal correlation time τ_s obtained from the extended Lipari and Szabo approach closely matches the center of the distribution of correlation times τ_0 . The extended Lipari and Szabo approach also contains an overall correlation time, whose optimal value was equal to $5 \text{ ns}\cdot\text{rd}^{-1}$ for the D2 domain. The interpretation of such an overall correlation time is difficult in the context of unfolded protein. It is located on the right side of the distribution, and appears necessary to compensate for the contribution of the slowest motions in the nanosecond time scale that are not taken into account by τ_s and S^2_s . The generalized order parameter S^2_f associated with fast internal correlation times is similar to the parameter A_{LD} . In addition to the analogy between τ_s and τ_0 , the evolution along the sequence of the D2 domain

Fig. 3. Comparison of the distributions of correlation times obtained for three residues of domain 2 of annexin I, Ala11, Ile34, and Asp60 using the three approaches described in the text: lorentzian distribution of correlation times (*upper panel*), extended model free (*middle panel*), and local model free (*lower panel*). A continuous line is used for the lorentzian distribution. Discrete values of correlation times are represented by vertical bars with a height corresponding to their associated order parameter.

of the order parameter, S^2_s matches the evolution of the width Δ of the distribution. Unexpectedly, the correlated evolution of S^2_s and Δ along the sequence suggests that the order parameter S^2_s accounts here for the dynamical heterogeneity of motions, rather than for amplitude of motions. This is highlighted by the higher values of S^2_s (correlated to higher values of Δ) found in the regions containing non-native structures around helix. In this case, this heterogeneity of motions seems related to a heterogeneity of distinct conformational states. Whether this observation can be generalized remains to be evaluated.

Taking into account the inherent flexibility of partially folded states, formalisms based on distributions of correlation times allow a suitable interpretation of relaxation data. The present analysis based on a lorentzian distribution of correlation times yields a clear picture of the dynamical behavior of the partially folded domain 2 of annexin I. It allows us to distinguish between the various types of residual structures found in this domain. It brings a more appropriate description of motions in unfolded states than the model-free approach considering discrete correlation times. Besides, it allows a reinterpretation of the parameters obtained from the extended model-free approach. It highlights the need to take into account the heterogeneity of conformational states in the nanosecond time scale for the interpretation of relaxation data obtained on partially folded states.

Materials and methods

Derivation of the spectral density function corresponding to a lorentzian distribution of correlation times

The general expression for the spectral density function in case of a distribution $F(\tau)$ of correlation time is:

$$J(\omega) = \int_0^{+\infty} F(\tau) J(\omega, \tau) d\tau \quad (1a)$$

$$\int_0^{+\infty} F(\tau) d\tau = 1 \quad (1b)$$

$$J(\omega, \tau) = \frac{2}{5} \frac{\tau}{1 + (\omega\tau)^2} \quad (1c)$$

We used here a lorentzian distribution characterized by:

$$F(\tau) = K \frac{\Delta}{\Delta^2 + (\tau - \tau_0)^2} \text{ for } 0 \leq \tau \leq \tau_{\max} \quad (2a)$$

$$F(\tau) = 0 \text{ for } \tau > \tau_{\max} \quad (2b)$$

K is a normalization constant such as to satisfy equation 1b, leading to:

$$\frac{1}{K} = \arctg \frac{\tau_{\max} - \tau_0}{\Delta} + \arctg \frac{\tau_0}{\Delta} \quad (2c)$$

The use of an upper limit $\tau_{\max}$ is imposed by the divergent behavior of the integral $J(0)$ when τ tends towards infinity. Equation (1a) then becomes:

$$J_{LD}(\omega) = \frac{1}{5D} \left\{ 2\tau_0[1 + S\omega^2] - 4K\Delta\omega\tau_0 \arctg(\omega\tau_{\max}) + K\Delta[1 - S\omega^2] \ln \frac{\Delta^2 + (\tau_{\max} - \tau_0)^2}{S(1 + \omega^2\tau_{\max}^2)} \right\} \\ S = (\Delta^2 + \tau_0^2), D = \{1 - 2(\Delta^2 - \tau_0^2)\omega^2 + S^2\omega^4\}. \quad (3)$$

A contribution of very fast internal motions in the picosecond range was also considered, which was assumed to be independent of the motions in the nanosecond range. The final expression of the spectral density function $J_{LD2}(\omega)$ is then:

$$J_{LD2}(\omega) = A_{LD} J_{LD}(\omega) + \frac{2}{5} (1 - A_{LD}) \frac{\tau_f}{1 + (\omega\tau_f)^2} \quad (4)$$

The second term in the right part of the equation was considered negligible, so that the contribution of fast libration motions was taken into account through the factor A_{LD} . This factor A_{LD} is analogous to the generalized order parameters S^2 or S^2_f used in the model-free approach.

Goodness of fits and uncertainties evaluation

$J(\omega)$ in equations 3–4 is a function of one to five unknown variables that can thus theoretically be evaluated from the nine independently measured relaxation parameters at the three fields. The additional exchange broadening contribution to transverse relaxation parameters $R_N(N_{x,y})$ was removed from the experimental values. No exchange contribution has thus to be taken into account in the further described fits.

Motion parameters were optimized for each residue for which the data could be collected from nonlinear least-squares minimization of the error function χ^2_{exp} using the Levenburg-Marquardt algorithm (Press et al. 1986) implemented in the MATLAB software.

$$\chi^2_{\text{exp}} = \sum_{\text{all frequencies}} \left[\left(\frac{R_N(N_{x,y})_{\text{calc}} - R_N(N_{x,y})_{\text{exp}}}{\delta R_N(N_{x,y})} \right)^2 + \left(\frac{R_N(N_z)_{\text{calc}} - R_N(N_z)_{\text{exp}}}{\delta R_N(N_z)} \right)^2 + \left(\frac{NOE_{\text{calc}} - NOE_{\text{exp}}}{\delta NOE} \right)^2 \right]$$

where $\delta(X)$ (X being $R_N[N_{x,y}]$, $R_N[N_z]$, or NOE) is the estimated error of the relaxation parameters. The subscripts exp and calc refer to experimental and back-calculated relaxation parameters, respectively. Initial values for parameters were obtained from a grid search minimization. The goodness of the fits and the uncertainties were obtained from statistical analysis (Palmer et al. 1991; Mandel et al. 1995). Five hundred simulated data sets were selected from a Gaussian noise distribution centered on the relaxation rates back calculated from the optimized parameters and corresponding to the previously estimated errors in $R_N(N_{x,y})$, $R_N(N_z)$, and NOEs. A model was considered satisfactory if the optimized χ^2_{exp} was within the 95% confidence limit obtained from the 500 Monte Carlo simulations, that is, $\chi^2_{\text{exp}} < \chi^2_{95\%}$. The uncertainties in the values of the model-free parameters are the standard deviation of the simulated parameters.

Acknowledgments

This work was supported by the Centre National de la Recherche Scientifique (CNRS). We thank F. Lauprêtre for enlightening discussions. The 800 MHz spectrometer was purchased with the financial help of the Conseil Régional d'Ile de France and of the Association pour la Recherche contre le Cancer.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Alexandrescu, A.T. and Shortle, D. 1994. Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease. *J. Mol. Biol.* **242**: 527–546.
- Brutscher, B., Bruschweiler, R., and Ernst, R.R. 1997. Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by ¹H, ¹³C, and ¹⁵N nuclear magnetic resonance spectroscopy. *Biochemistry* **36**: 13043–13053.
- Buck, M., Schwalbe, H., and Dobson, C.M. 1996. Main-chain dynamics of a partially folded protein: ¹⁵N NMR relaxation measurements of hen egg white lysozyme denatured in trifluoroethanol. *J. Mol. Biol.* **257**: 669–683.
- Buevich, A.V. and Baum, J. 1999. Dynamics of unfolded proteins: Incorporation of distributions of correlation times in the model free analysis of NMR relaxation data. *J. Am. Chem. Soc.* **121**: 8671–8672.
- Buevich, A.V., Shinde, U.P., Inouye, M., and Baum, J. 2001. Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. *J. Biomol. NMR* **20**: 233–249.
- Clare, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., and Gronenborn, A.M. 1990. Deviation from the simple two-parameters model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* **112**: 4989–4991.
- Cordier-Ochsenbein, F., Guerois, R., Baleux, F., Huynh-Dinh, T., Chaffotte, A., Neumann, J.M., and Sanson, A. 1996. Folding properties of an annexin I domain: A ¹H-¹⁵N NMR and CD study. *Biochemistry* **35**: 10347–10357.
- Cordier-Ochsenbein, F., Guerois, R., Baleux, F., Huynh-Dinh, T., Lirsac, P.N., Russo-Marie, F., Neumann, J.M., and Sanson, A. 1998. Exploring the folding pathways of annexin I, a multidomain protein. I. Non-native structures stabilize the partially folded state of the isolated domain 2 of annexin I. *J. Mol. Biol.* **279**: 1163–1175.
- Dejean de la Batie, R., Lauprêtre, F., and Monnerie, L. 1988. Carbon-13 NMR investigation of local dynamics in bulk polymers at temperatures well above the glass transition temperature. I. Poly(vinyl methyl ether). *Macromolecules* **21**: 2045–2052.
- Desvaux, H. and Berthault, P. 1999. Study of dynamic processes in liquids using off-resonance rf irradiation processes. *Prog. NMR Spectrosc.* **35**: 295–340.
- Dyson, H.J. and Wright, P.E. 1998. Equilibrium NMR studies of unfolded and partially folded proteins. *Nat. Struct. Biol. Suppl.* **5**: 499–503.
- . 2001. Nuclear magnetic resonance methods for elucidation of structures and dynamics in disordered states. *Methods Enzymol.* **339**: 258–270.
- Eliez, D., Chung, J., Dyson, H.J., and Wright, P.E. 2000. Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry* **39**: 2894–2901.
- Eliez, D., Yao, J., Dyson, H.J., and Wright, P.E. 1998. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* **5**: 148–155.
- Farrow, N.A., Zhang, O., Forman-Kay, J.D., and Kay, L.E. 1997. Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry* **36**: 2390–2402.
- Frank, M.K., Clare, G.M., and Gronenborn, A.M. 1995. Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. *Protein Sci.* **4**: 2605–2615.
- Heatley, F. 1979. Nuclear magnetic relaxation of synthetic polymers in dilute solutions. *Prog. NMR Spectrosc.* **13**: 47–85.
- . 1986. Nuclear magnetic relaxation and models for backbone motions of macromolecules in solution. *Annu. Rep. NMR Spectrosc.* **17**: 179–230.
- Huber, R., Berendes, R., Burger, A., Schneider, M., Karshikov, A., Luecke, H., Romisch, J., and Paques, E. 1992. Crystal and molecular structure of human annexin V after refinement. Implications for structure, membrane binding and ion channel formation of the annexin family of proteins. *J. Mol. Biol.* **223**: 683–704.
- Kataoka, M., Kuwajima, K., Tokunaga, F., and Goto, Y. 1997. Structural characterization of the molten globule of alpha-lactalbumin by solution X-ray scattering. *Protein Sci.* **6**: 422–430.
- Kataoka, M., Nishii, I., Fujisawa, T., Ueki, T., Tokunaga, F., and Goto, Y. 1995. Structural characterization of the molten globule and native states of apomyoglobin by solution X-ray scattering. *J. Mol. Biol.* **249**: 215–228.
- Korzhnev, D.M., Orekhov, V.Y., and Arseniev, A.S. 1999a. The effect of helix-coil transition on backbone ¹⁵N NMR relaxation of isolated transmembrane segment (1–36)-bacteriorhodopsin. *J. Biomol. NMR* **14**: 357–368.
- Korzhnev, D.M., Orekhov, V.Y., Arseniev, A.S., Gratias, R., and Kessler, H. 1999b. Mechanism of the unfolding of transmembrane α -helical segment (1–36)-bacteriorhodopsin studied by dynamics simulations. *J. Phys. Chem. B* **103**: 7036–7043.
- Kraulis, P.J. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**: 946–950.
- Kroenke, C., Loria, J., Lee, L., Rance, M., and Palmer, A. 1998. Longitudinal and transverse ¹H-¹⁵N dipolar/¹⁵N chemical shift anisotropy relaxation interference: Unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. *J. Am. Chem. Soc.* **120**: 7905–7915.
- Lipari, G. and Szabo, A. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecule. I. Theory and range of validity. *J. Am. Chem. Soc.* **104**: 4546–4559.
- Mandel, A.M., Akke, M., and Palmer, A.G. 1995. Backbone dynamics of *Escherichia coli* ribonuclease HI: Correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246**: 144–163.
- Meekhof, A.E. and Freund, S.M. 1999. Probing residual structure and backbone dynamics on the milli- to picosecond timescale in a urea-denatured fibronectin type III domain. *J. Mol. Biol.* **286**: 579–592.
- Mulder, F.A.A., de Graaf, R.A., Kaptein, R., and Boelens, R. 1998. An off-resonance rotating frame relaxation experiment for the investigation of macromolecular dynamics using adiabatic rotations. *J. Magn. Reson.* **131**: 351–357.
- Ochsenbein, F., Guerois, R., Neumann, J.M., Sanson, A., Guittet, E., and van Heijenoort, C. 2001. ¹⁵N NMR relaxation as a probe for helical intrinsic propensity: The case of the unfolded D2 domain of annexin I. *J. Biomol. NMR* **19**: 3–18.
- Orekhov, V.Y., Korzhnev, D.M., Pervushin, K.V., Hoffmann, E., and Arseniev, A.S. 1999. Sampling of protein dynamics in nanosecond time scale by ¹⁵N NMR relaxation and self-diffusion measurements. *J. Biomol. Struct. Dyn.* **17**: 157–174.
- Palmer, A.G.I., Rance, M., and Wright, P.E. 1991. Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance ¹³C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.* **113**: 4371–4380.
- Phan, I.Q.H., Boyd, J., and Campbell, I.D. 1996. Dynamic studies of a fibronectin type I module pair at three frequencies: Anisotropic modelling and direct determination of conformational exchange. *J. Biomol. NMR* **8**: 369–378.
- Press, W.H., Flannery, B.P., Teutolski, S.A., and Vetterling, W.T. 1986. *Numerical recipes, the art of scientific computing*. Cambridge University Press, Cambridge.
- Schwalbe, H., Fiebig, K.M., Buck, M., Jones, J.A., Grimshaw, S.B., Spencer, A., Glaser, S.J., Smith, L.J., and Dobson, C.M. 1997. Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* **36**: 8977–8991.
- van Heijenoort, C., Penin, F., and Guittet, E. 1998. Dynamics of the DNA binding domain of the fructose repressor from the analysis of linear correlations between the ¹⁵N-¹H bond spectral densities obtained by nuclear magnetic resonance spectroscopy. *Biochemistry* **37**: 5060–5073.
- Weng, X., Luecke, H., Song, I.S., Kang, D.S., Kim, S.H., and Huber, R. 1993. Crystal structure of human annexin I at 2.5 Å resolution. *Protein Sci.* **2**: 448–458.