J Biomol NMR (2008) 42:163–177
DOI 10.1007/s10858-008-9280-0

ARTICLE

Model-independent interpretation of NMR relaxation data for unfolded proteins: the acid-denatured state of ACBP

Kristofer Modig · Flemming M. Poulsen

Received: 13 March 2008 / Accepted: 10 July 2008 / Published online: 11 October 2008

Abstract We have investigated the acid-unfolded state of acyl-coenzyme A binding protein (ACBP) using ^{15}N laboratory frame nuclear magnetic resonance (NMR) relaxation experiments at three magnetic field strengths. The data have been analyzed using standard model-free fitting and models involving distribution of correlation times. In particular, a model-independent method of analysis that does not assume any analytical form for the correlation time distribution is proposed. This method explains correlations between model-free parameters and the analytical distribution parameters found by other authors. The analysis also shows that the relaxation data are consistent with and complementary to information obtained from other parameters, especially secondary chemical shifts and residual dipolar couplings, and strengthens the conclusions of previous observations that three out of the four regions that form helices in the native structure appear to contain residual secondary structure also in the acid-denatured state.

Keywords ACBP · Protein folding · Acid-denaturation · NMR · Relaxation · Model-free

K. Modig (✉) · F. M. Poulsen
Structural Biology and NMR Laboratory, Department of Molecular Biology, University of Copenhagen, Copenhagen Biocenter, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark
e-mail: kristofer.modig@bpc.lu.se

Present Address:
K. Modig
Department of Biophysical Chemistry, Center for Molecular Protein Science, Lund University, Box 124, 22100 Lund, Sweden

Abbreviations

LD	Lorentzian distribution
MF	Model-free
MIC	Model-independent correlation time distribution
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser effect
PRE	Paramagnetic relaxation enhancement
RDC	Residual dipolar coupling
SDF	Spectral density function
SDM	Spectral density mapping
TCF	Time-correlation function

Introduction

The denatured state of a protein is the starting point in the folding reaction that passes over a series of transient conformations that may resemble the folded state (Shortle 1996; Teilum et al. 2000; Shortle and Ackerman 2001; Choy et al. 2002; Klein-Seetharaman et al. 2002; Millet et al. 2002; Teilum et al. 2002; Krieger et al. 2003; Kumar et al. 2003; Modig et al. 2003; Sanchez and Kiefhaber 2003; Fieber et al. 2004; Lindorff-Larsen et al. 2004; Halle et al. 2005; Krieger et al. 2005; Kristjansdottir et al. 2005; Lindorff-Larsen et al. 2005; Teilum et al. 2006). To investigate denatured protein states in search for the folding-inducing conformations is a challenging task, because the important conformations in the denatured ensemble may be sparsely populated and difficult to detect. Nuclear magnetic resonance (NMR) is undoubtedly one of the most useful methods in the study of denatured states, because of the multitude of parameters that can be extracted. These parameters contain both structural and dynamical information. During the past decade, theoretical, methodological and computational advances have made it possible

to extract useful information also on the elusive non-native states by NMR methods. For example, paramagnetic relaxation enhancement (PRE) experiments have been used to probe transient long-range interactions (Teilum et al. 2002; Lindorff-Larsen et al. 2004; Kristjansdottir et al. 2005), residual dipolar couplings (RDC) have been interpreted in terms of residual structure (Shortle and Ackerman 2001; Ackerman and Shortle 2002; Alexandrescu and Kammerer 2003; Fieber et al. 2004; Sallum et al. 2005), and the folding reaction has been studied by relaxation methods sensitive to dynamics at the μs-ms timescale (Teilum et al. 2006).

NMR relaxation rates, such as the longitudinal R_1, the transverse R_2 and the heteronuclear nuclear Overhauser effect (NOE), carry unique information on the dynamics within the ensemble of denatured molecules. These rates are sensitive to motions on the ps–ns time-scale, but are difficult to interpret for flexible protein chains. Nevertheless, they have been used extensively and interpreted in various ways. Most of the previous studies have employed reduced spectral density mapping (SDM) (Farrow et al. 1995) or the so-called model-free (MF) approach (Halle and Wennerström 1981; Lipari and Szabo 1982; Clore et al. 1990) in order to probe residual, predominantly helical, structure content (Alexandrescu and Shortle 1994; Ochsenbein et al. 2001; Choy and Kay 2003; Cao et al. 2004; Zhang et al. 2005; Shojania and O'Neil 2006). However, spectral density mapping gives only the value of the spectral density at certain frequencies and might be difficult to interpret, while the model-free approach has some underlying assumptions that question its validity for very flexible protein states (*vide infra*). Therefore, other schemes of interpretations have been applied, for example polymer models involving segmental clusters with certain persistence lengths (Schwalbe et al. 1997; Klein-Seetharaman et al. 2002; Schwarzinger et al. 2002), analytical correlation time distributions (Buevich and Baum 1999; Ochsenbein et al. 2002) and molecular dynamics simulations (Prompers and Brüschweiler 2002).

Here, we use laboratory frame relaxation data for flexible peptide chains to asses the residue-specific correlation time distribution by extracting its fundamental statistical quantities (mean, standard deviation and skewness) without any assumption on the analytical form of the distribution. This model-independent correlation time distribution (MIC) analysis was previously introduced in a different context (Halle et al. 1998), but is adapted here to suit the analysis of high-resolution NMR relaxation data. Furthermore, the MIC analysis allows the calculation of fundamental properties, such as the half-time, of the fundamental time-correlation function (TCF) that induces the relaxation. Finally, we compare the parameters obtained in the MIC and MF analyses with those from models involving analytical distribution functions.

The MIC analysis is applied to the acid-denatured state of bovine acyl-coenzyme A binding protein (ACBP), an 86-residue 4-helix bundle. Its folding behavior and denatured states (at low pH and guanidine hydrochloride) have been studied extensively (Kragelund et al. 1995; Kragelund et al. 1999; Teilum et al. 2000; Thomsen et al. 2002; Teilum et al. 2006). Specifically, paramagnetic relaxation enhancement experiments have clearly shown that the ensembles within the denatured states are quite similar under the two denaturing conditions (Teilum et al. 2002; Lindorff-Larsen et al. 2004; Kristjansdottir et al. 2005). These studies showed that helices 2 and 3 make transient contacts in the denatured states, and helices 2 and 4 make additional contact under mildly denaturing conditions. Residual dipolar couplings and secondary chemical shift patterns clearly identify the four regions that are helical in the native protein as having some residual helical structure (Teilum et al. 2002; Thomsen et al. 2002; Fieber et al. 2004; Modig et al. 2007). The present analysis shows that the correlation time distribution is more skewed towards longer times in helix A3 and A4 than in the rest of the protein. The half-time profile shows significant maxima at the positions of helices A2, A3 and A4, reinforcing the view that these segments form clusters that tumble more slowly than the rest of the protein.

Theory

The relaxation rates for the ^{1}H–^{15}N spin pair are (Abragam 1961)

$$\begin{aligned} R_1 &= \frac{d^2}{4}[3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)] \\ &\quad + c^2 J(\omega_N) \\ R_2 &= \frac{d^2}{8}[4J(0) + 3J(\omega_N) + 6J(\omega_H) + J(\omega_H - \omega_N) \\ &\quad + 6J(\omega_H + \omega_N)] + \frac{c^2}{6}[J(0) + 3J(\omega_N)] + R_{ex} \\ \sigma &= \frac{d^2}{4}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \end{aligned} \quad (1)$$

and the heteronuclear steady-state *NOE* is $NOE = 1 + (\gamma_H/\gamma_N)\sigma/R_1$. $d^2 = (\mu_0 h \gamma_N \gamma_H / 8\pi^2)^2 \langle r^{-3} \rangle^2$ and $c^2 = (\Delta\sigma\omega_N)^2/3$, h is Plancks's constant, μ_0 is the vacuum permeability constant, γ_is are the gyromagnetic ratios, $r = 1.04$ Å is the length of the N–H vector, and $\Delta\sigma = -163$ ppm is the shielding anisotropy (Palmer 2004). As long as in the protein dynamics fall the regime where standard Bloch–Wangsness–Redfield perturbation theory applies, these equations are generally valid and the main challenge is to formulate expressions for the spectral density $J(\omega)$. The spectral density function (SDF) is the

real part of the Fourier transform of the time correlation function (TCF):

$$J(\omega) = \frac{1}{2}\mathrm{Re}\int_{-\infty}^{\infty} C(\tau)\exp(-i\omega\tau)\,\mathrm{d}\tau \tag{2}$$

In general, the TCF is modeled as a sum of exponential functions, and, hence, the SDF is a sum of Lorentzian components,

$$J(\omega) = C(0)\sum_{k=1}^{N} c_k \frac{\tau_k}{1+(\omega\tau_k)^2} \tag{3}$$

where the number of Lorentzians N, the coefficients c_k and the correlation times τ_k, vary according to the chosen model. These expressions are general and hold for any stationary Markov process obeying detailed balance (van Kampen 1981), for example jump processes, as well as both free and restricted rotational diffusion (Favro 1960; Ferrarini et al. 1994). Specifically, in the standard model-free approach, valid in the case of motions on two different time-scales (Halle and Wennerström 1981) or in the case of two statistically uncorrelated motions (Lipari and Szabo 1982)

$$\begin{aligned} &c_1 = S^2; c_2 = 1 - S^2 \\ &\tau_1 = \tau_\mathrm{m}; \tau_2 = \tau_\mathrm{e} \end{aligned} \tag{4}$$

where τ_m is the overall 2nd rank correlation time and τ_e is the correlation time for the restricted internal motions of the N–H vector within the diffusion frame of the protein, given by $1/\tau_\mathrm{e} = 1/\tau_\mathrm{m} + 1/\tau_\mathrm{f}$. S is the order parameter that describes the restriction on the internal motions. In the limit $S^2 = 1$, the N–H vector is completely fixed and in the (unphysical) limit $S^2 = 0$, the N–H vector can have any direction within the diffusion frame. Here, the root-mean-square fluctuations in the correlation function is $C(0) = 2/5$, since the lattice parameters $\langle r^{-3}\rangle^2$ and $\Delta\sigma$ are placed in Eq. 1 and are assumed to be known. Extensions to the model-free spectral density, including anisotropic tumbling or motions on two internal time-scales, exist (Woessner 1962; Clore et al. 1990; Palmer 2004). For example, if there are motions on two internal time-scales, then

$$\begin{aligned} &c_1 = S_f^2 S_s^2; c_2 = S_f^2\left(1 - S_s^2\right); c_3 = 1 - S_f^2 \\ &\tau_1 = \tau_\mathrm{m}; \tau_2 = \tau_\mathrm{s}; \tau_3 = \tau_\mathrm{e} \approx 0 \end{aligned} \tag{5}$$

Another common case is that of an anisotropic diffusion tensor, for which $J(\omega)$ is a sum similar to Eq. 3 with at least three terms (Woessner 1962).

For partially and fully denatured proteins, one cannot define an overall diffusion frame and one often resorts to defining a residue-specific correlation time τ_loc in place of the global τ_m value. Even though this is correct in principle, the assumptions of time-scale or statistical separation underlying Eq. 4 may not hold. Another problem arises because the distribution $f(\tau_\mathrm{loc})$ can be expected to be rather broad for a denatured protein. There are two reasons for this. First, the ensemble of conformational states is large for a denatured protein. It consists of both compact and extended conformers, which have different overall tumbling times. Second, a denatured protein will have many motional modes (Ribeiro et al. 1980; Prompers and Brüschweiler 2002). For example, in one peptide segment all residues will move in a correlated fashion, but even segments distant in sequence might move correlated because they (temporarily) interact mutually. Each of these modes has its own correlation time. Both these aspects make the specification of a single local correlation time per residue dubious. One simple solution to this is to specify some analytical distribution function $f(\tau_\mathrm{C})$ for the correlation times, and then apply the spectral density function

$$J(\omega) = C(0)\int_0^{\infty} f(\tau_\mathrm{C})\frac{\tau_\mathrm{C}}{1+(\omega\tau_\mathrm{C})^2}\,\mathrm{d}\tau_\mathrm{C} \tag{6}$$

where we introduce a correlation time τ_C instead of τ_loc. Hereafter, τ_loc will be reserved to the context of the local model-free model. At least two such analytical distribution functions have been proposed in the literature. The first is a distribution introduced by Cole and Cole for use in the analysis of dielectric relaxation dispersions of viscous materials (Cole and Cole 1941), and later introduced in the context of NMR relaxation of denatured proteins (Buevich and Baum 1999; Buevich et al. 2001). The Cole–Cole distribution has the fundamental problem of diverging when $\omega \to 0$, and therefore not being optimally suited for the analysis of R_2 relaxation rates that contain the spectral density $J(0)$. The second is the Lorentzian distribution (LD) (Ochsenbein et al. 2002), characterized by a the central correlation time τ_0 and the width Δ:

$$f_\mathrm{LD}(\tau_\mathrm{C}) = K\frac{\Delta}{\Delta^2 + (\tau_\mathrm{C} - \tau_0)^2},\ \text{for } 0 \le \tau_\mathrm{C} \le \tau_\mathrm{max} \tag{7a}$$

$$f_\mathrm{LD}(\tau_\mathrm{C}) = 0,\ \text{for } \tau_\mathrm{C} > \tau_\mathrm{max} \tag{7b}$$

$$\frac{1}{K} = \arctan\frac{\tau_\mathrm{max} - \tau_0}{\Delta} + \arctan\frac{\tau_0}{\Delta} \tag{7c}$$

The introduction of τ_max is needed to make the integral in Eq. 6 converge. A second correlation time can be introduced in analogy with Eq. 4, so that the final spectral density is

$$J(\omega) = \frac{2}{5}\left[S_\mathrm{LD}^2\int_0^{\infty} f_\mathrm{LD}(\tau_\mathrm{C})\frac{\tau_\mathrm{C}}{1+(\omega\tau_\mathrm{C})^2}\,\mathrm{d}\tau_\mathrm{C} + \frac{\left(1 - S_\mathrm{LD}^2\right)\tau_\mathrm{e}}{1+(\omega\tau_\mathrm{e})^2}\right] \tag{8}$$

Both these analytical distributions are formally symmetric. However, since τ_C cannot be negative, $f_\mathrm{LD}(\tau_\mathrm{C})$ has a long

slowly decaying tail towards τ_{max}, but ends abruptly with a non-zero value at the other limit $\tau_C = 0$. Thus, if τ_0 is not much larger than Δ, $f_{LD}(\tau_C)$ is always a skewed distribution with positive skewness (see below). The introduction of the fast correlation time can thus be regarded as a correction for the case when the true distribution is shifted in the other direction, i.e. has a zero or negative skewness.

Model-independent correlation time distribution analysis

Another compelling method is to model the shape of the spectral density function and then to draw some conclusions on the underlying $f(\tau_C)$ from the shape only. Such a model-independent correlation time distribution (MIC) analysis has previously been done in the analysis of the stretched magnetic relaxation dispersion curves often obtained when monitoring the frequency dependence of water proton relaxation rates of protein solutions (Halle et al. 1998). By setting $\omega = 0$ in Eq. 6 one may define the average correlation time as

$$\langle \tau_C \rangle = \frac{J(0)}{C(0)} \tag{9}$$

Now, it is well known that a distribution like that of Eq. 6 can be approximated by a sum of exponentials with only a few terms, like in Eq. 3. It is then possible to obtain some key statistical properties of $f(\tau_C)$ from the coefficients c_k and correlation times τ_k directly. If many relaxation rates are obtained over the full frequency range (as in the original application using the fast field-cycling technique), it is thus possible to obtain the "true" mean-square fluctuation $C'(0)$ as well as $f(\tau_C)$. In this case, one can use the relation $C'(0) = 2/\pi \int_0^\omega J(\omega)\,d\omega$ (Halle et al. 1998), in which we replace $J(\omega)$ with Eq. 3. Then, $C'(0) = C(0)\sum_k c_k$. In high-resolution NMR, we can only obtain ^{15}N relaxation rates at a small number of magnetic field strengths, and we choose to use the common value $C(0) = 2/5$ and force $\sum_k c_k = 1$, so that $C'(0) = C(0)$. We can now extract some key quantities. The mean of the correlation times in the distribution is obtained from the combination of Eqs. 9 and 3:

$$\langle \tau_C \rangle = \sum\nolimits_k c_k \tau_k \tag{10}$$

The central moments of $f(\tau_C)$ are (Weisstein 2006a)

$$\mu_n = \int_0^\infty f(\tau_C)(\tau_C - \langle \tau_C \rangle)^n \, d\tau_C \approx \sum\nolimits_k c_k (\tau_k - \langle \tau_C \rangle)^n \tag{11}$$

These are used to calculate the standard deviation σ_τ and skewness γ_τ of the distribution as (Weisstein 2006b)

$$\sigma_\tau = \sqrt{\mu_2} \tag{12a}$$

$$\gamma_\tau = \frac{\mu_3}{\mu_2^{3/2}} \tag{12b}$$

The skewness is negative when the left tail (small τ_C values) of $f(\tau_C)$ is more pronounced than the right tail. The higher moments ($n > 2$) are notoriously difficult to calculate and require a larger number of Lorentzian terms in the approximation of the true sum. However, it is possible to estimate some very fundamental properties of the correlation function $C(\tau)$. The first is the time $\tau_{1/2}$ at which the correlation time has decayed to half its initial value. Since the spectral density is a sum of Lorentzian functions, the time correlation function is a sum of exponential functions. Thus, $\tau_{1/2}$ is calculated from the implicit relation

$$\frac{1}{2} = \sum_k c_k \exp\left(-\tau_{1/2}/\tau_k\right) \tag{13}$$

The second fundamental property is how much of the TCF that remains for times larger than a certain cut off, for example $\tau > \tau_{1/2}$. The most simple measure is probably the ratio between the integral of the tail above $\tau_{1/2}$ and the full TCF integral, $\int_{\tau_{1/2}}^\infty C(\tau)\,d\tau / \int_0^\infty C(\tau)\,d\tau$. The ratio is

$$\lambda = \frac{\sum_k \tau_k c_k \exp\left(-\tau_{1/2}/\tau_k\right)}{\sum_k c_k \tau_k} \tag{14}$$

If the short correlation times in $f(\tau_C)$ dominate, $C(\tau)$ decays rapidly to half its initial value (i.e. $\tau_{1/2}$ is small) but the tail is long and $\lambda \to 1$. Conversely, if the long correlation times dominate, there is only a small initial fall of $C(\tau)$, $\tau_{1/2}$ is large and $\lambda \to 0.5$, the value for a correlation function with only one exponential term. For example, for the Lipari–Szabo spectral density with $\tau_m = 4$ ns and $\tau_e = 50$ ps, $[\tau_{1/2}, \lambda]$ equals [1.88 ns, 0.62] and [49 ps, 0.96] for $S^2 = 0.8$ and 0.2, respectively. In order to test whether a small number of Lorentzians can represent a distribution $f(\tau)$, we generated synthetic data sets with the help of Eq. 6. 300 τ_k values were linearly spaced between 0 and 10 ns and the coefficients c_k followed the distribution

$$c_k \sim \left[1 - \exp\left(-\frac{\tau_k^2}{\tau_{cut}^2}\right)\right]\left[\frac{p}{\sigma_1}\exp\left(-\frac{(\tau_k - \bar{\tau}_1)^2}{2\sigma_1}\right) + \frac{(1-p)}{\sigma_2}\exp\left(-\frac{(\tau_k - \bar{\tau}_2)^2}{2\sigma_2}\right)\right] \tag{15}$$

The first factor, with $\tau_{cut} = 0.03$ ns, ensures that the distribution goes smoothly to zero when $\tau_k \to 0$. The coefficients were normalized so that $\sum c_k = 1$. We then used the sets $\{\tau_k, c_k\}$ to calculate the spectral densities $J(0)$, $J(50\ \text{MHz})$, $J(60\ \text{MHz})$, $J(80\ \text{MHz})$, $J(500\ \text{MHz})$,

J(600 MHz) and J(800 MHz). The parameters used were all possible combinations of $\bar{\tau}_1 = 0.05$ ns, $\bar{\tau}_2 = [1, 2, 3]$ ns, $\sigma_1 = 0.1$ ns and $\sigma_2 = [0.1, 1, 2]$ ns. p was set to 10 values between 0.3 and 0.8. Random noise with a standard deviation of 0.01 $J(0)$ was added to the data points. Each synthetic data set was then fitted to Eq. 3 with only two terms and $\langle\tau_C\rangle$,σ_τ, γ_τ, $\tau_{1/2}$ and λ were calculated for both the simulated and fitted $\{c_k,\tau_k\}$ sets. Finally, only the fits where the shortest correlation time was shorter than 200 ps were retained, in order to fall into the regime of the experimental data. The results are shown in Fig. 1.

It should be noted that the skewness is very unreliable if the separation between the fitted correlation times is small (not shown in Fig. 1), whereas both $\langle\tau_C\rangle$ and σ_τ estimate their true values well over the full range. However, in the range spanned by the experimental data, all parameters correlate reasonably well to their true values. In the case of $\langle\tau_C\rangle$ and σ_τ the agreements are very good, but the spread in the skewness is larger and the fits underestimate the skewness by 0.5 on average, although it appears that the true values are never smaller than the fitted ones. Thus, the fitted skewness can be regarded as a lowest limit. The lower precision in the skewness is not surprising, since the skewness involves calculation of the third moment of the distribution. However, when considered on average over a protein segment the skewness can be expected to be a good indicator on the trends of the shape of the underlying distribution along an amino acid chain.

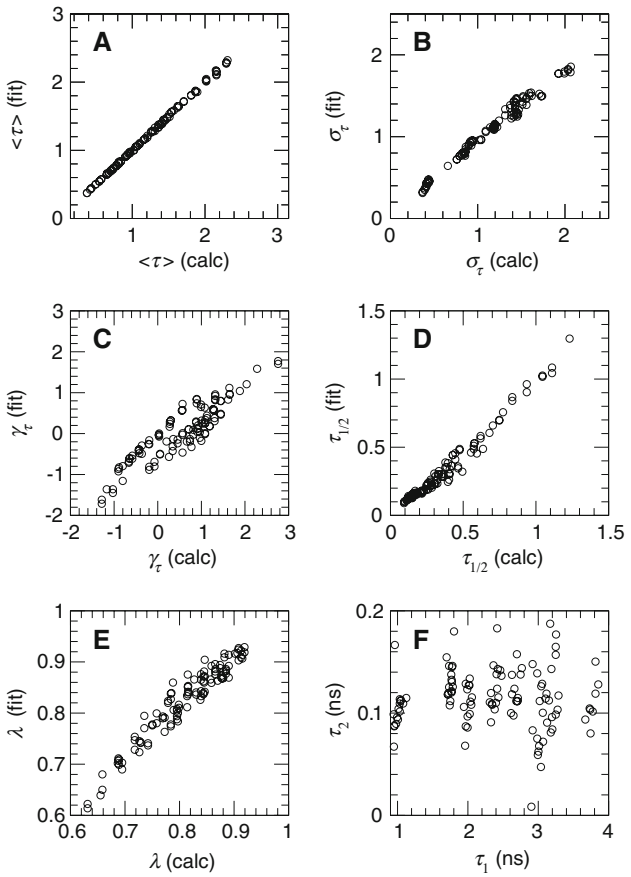

Fig. 1 The result of a 2-Lorentzian fit to data generated using 300 Lorentzians with $\{c_k, \tau_k\}$ generated as described in the text. Panels (**a**–**c**) show the average ($\langle\tau_C\rangle$), the standard deviation (σ_τ) and skewness (γ_τ) of $f(\tau_C)$. Panels (**d**–**e**) show the half-time ($\tau_{1/2}$) and tail integral ratio λ of the TCF, respectively. Panel (**f**) shows the distribution of the fitted τ_1 and τ_2 values of the two Lorentzian terms

It should be emphasized that this TCF sculpting is model-independent, except in the choice of the rigid lattice parameters $\langle r^{-3}\rangle^2$ and $\Delta\sigma$. Especially the CSA might very well vary from residue to residue, but that variation is manifested predominantly at the higher magnetic fields. The distribution of τ_k can have two origins. In the first, the protein ensemble is consisting of separate conformers that interconvert on a much longer time-scale than the tumbling time of the individual members. Since the interconversion is too slow to contribute to the relaxation, the observed distribution is the distribution of tumbling-times among the conformers. In the second view, the interconversion occurs on the same or shorter time-scales as the tumbling. Then, the distribution is the distribution of the correlation times of the motional modes of correlated motion within the ensemble (Ribeiro et al. 1980; Prompers and Brüschweiler 2002). Unfortunately, it is not a priori possible to distinguish between the two, and the most probable scenario is a mix of both. Some ensemble conformations exist for long times (longer than 10 ns or so), whereas other interconvert on shorter time-scales.

In order to fit experimental data, one might either first perform a spectral density mapping (Farrow et al. 1995) and then directly fit the $J(\omega)$ values, or first perform a standard local model-free analysis (Mandel et al. 1995) and reinterpret the parameters according to Eqs. 4 or 5. Both procedures have been used with very similar results; hence, only the model-free path is presented here. It should be pointed out that the sum of Lorentzians in Eq. 3 serves as an interpolation of the true underlying distribution. Therefore, the local model-free method should be regarded only as a convenient way of obtaining the parameters $\{c_k,\tau_k\}$ and, if needed, one does not have to consider the physical sense of, for example, using the extended model-free expressions (Eq. 5) with the global τ_m replaced by a local τ_{loc}.

Apparent correlation between MF and distribution parameters

Since the model-free SDF can be used to asses the coefficients and correlation times in Eq. 3, it is not surprising that correlations between the distribution properties and the model-free parameters are seen (Buevich et al. 2001; Ochsenbein et al. 2002). In order to clarify these

Table 1 Expressions or leading terms for the mean, variance and skewness of the distribution $f(\tau_C)$ in terms of model-free parameters

Property	Standard local MF	Extended MF
$\langle\tau_C\rangle$	$(\tau_{loc} - \tau_e)S^2 + \tau_e$	$S_f^2 S_s^2(\tau_m - \tau_s) + S_f^2\tau_s$
σ_τ^2	$S^2(1 - S^2)(\tau_{loc} - \tau_e)^2$	$\tau_m^2 S_f^2 S_s^2\left(1 - S_f^2 S_s^2\right) + K^a$
γ_τ	$(1 - 2S^2)/\sqrt{S^2(1 - S^2)}$	ND[b]

[a] Only the leading term involving τ_m^2 is given

[b] γ_τ from the extended MF parameters is a lengthy expression containing all the parameters and is not considered here

correlations, we insert the parameters of Eqs. 4 and 5 into Eqs. 10 and 11. The full expressions or their leading terms are shown in Table 1. In the case of a local model-free fit, the mean is dominated by the product $\tau_{loc}S^2$ (if τ_e is small) and the standard deviation of $\tau_{loc}\sqrt{S^2(1 - S^2)}$. One should note that $S^2(1 - S^2)$ is symmetric about its maximum value at $S^2 = 0.5$. Therefore, the sequence profile of σ_τ/τ_{loc} is correlated with S^2 if $S^2 < 0.5$ and anti-correlated if $S^2 > 0.5$. In the case of an extended model-free fit, with a global τ_m and a sequence-dependent τ_s, the variations in σ_τ stem only from the product $S_f^2 S_s^2\left(1 - S_f^2 S_s^2\right)$ (to first order). If the S_f^2 profile is fairly flat, the variations are almost only due to variations in S_s^2 and vice versa. Again, σ_τ is symmetric around $S_f^2 S_s^2 = 0.5$. For the mean, we focus on two cases. If τ_s is much smaller than τ_m, then the variations in the mean along the sequence stem from the product $S_f^2 S_s^2$. On the other hand, if S_s^2 is fairly small compared to S_f^2, then the variations in the mean can be expected to follow the $S_f^2\tau_s$ profile. Finally, the skewness is a very simple expression involving only S^2, in terms of the local MF parameters. It varies from 2.5 to −2.5 in the interval $0.1 \leq S^2 \leq 0.9$ and diverges towards plus and minus infinity in the limits $S^2 \to 0$ and $S^2 \to 1$, respectively. Thus, the skewness profile will be closely correlated to the S^2 profile.

Materials and methods

Sample preparation and NMR experiments

^{15}N-labeled bovine wild-type ACBP was expressed and purified as described (Mandrup et al. 1991). The protein was dissolved in H_2O, containing 7.5% D_2O and the pH was adjusted to 2.3 (uncorrected for isotope effect) by adding small amounts of HCl. Since acid-unfolded ACBP forms dimers at higher concentrations, which in turn induce helix formation in helix A4 (Fieber et al. 2005), a low final protein concentration of approximately 50 μM was used. The NMR experiments were carried out at 27°C on Varian INOVA spectrometers operating at 499.86, 599.89 and 799.85 MHz. The temperature was calibrated using a standard sample of methanol. The pulse sequences used to determine ^{15}N R_1, R_2 and ^{15}N–^{1}H NOEs were those previously described (Farrow et al. 1994). The spectra were acquired with 128 complex data points in t_1, and the ^{15}N sweep-width set to 25 ppm at the 500 and 600 MHz spectrometers and to 21 ppm at the 800 MHz spectrometer. About 32 scans were acquired per free induction decay. ^{15}N decoupling during acquisition was done using GARP (Shaka et al. 1985). The relaxation delay of 10 s was used in the NOE experiments. Proton saturation was achieved with 120° proton pulses spaced at 5 ms intervals for 5 s and the saturated and unsaturated experiments were acquired in an interleaved manner. The processing was carried out using the nmrPipe software (Delaglio et al. 1995). The water peak was suppressed using the nmrPipe POLY function on the free-induction decay. A slightly shifted cosine bell window function was applied to both dimensions, and the spectra were extensively zero-filled to 1024 × 8192 complex data points. After Fourier transformation the spectra were base-line corrected in the proton dimension. Sparky (Goddard and Kneller 2005) was used for visualization and assignment of the spectra. The peak intensities, I, were determined by adding 15 points around each peak maxima using the program seriesTab in the nmrPipe package. Finally, the relaxation decays were fitted using in-house written scripts based on the program Gnuplot. A 2-parameter equation was used to fit both R_1 and R_2. The standard deviations of the NOEs were determined from the noise level of the saturated and unsaturated spectra as $\sigma_{NOE} = I_s/I_u\left[(15\sigma_s/I_s)^2 + (15\sigma_u/I_u)^2\right]^{1/2}$, where σ_s and σ_u are the single-point noise levels. For the R_1 and R_2 experiments, duplicate measurements were done for the shortest relaxation delay, and the standard deviation was calculated as $\sigma_{R_1,R_2} = \left[1/(2N)\sum_{j=1}^{N}\left(I_{j,1} - I_{j,2}\right)^2\right]^{1/2}$, where N is the number of peaks. This expression is equivalent to the pooled standard deviation between N sets of duplicate measurements.

Spectral density mapping and model selection

The reduced spectral density mapping (SDM) was performed as described by Farrow et al. (1995) but with the modification described in the Appendix. The fit was performed twice for each residue with and without the R_{ex} term. Thus, the presence of a chemical exchange contribution could be judged from the F-test (see below). The SDM analysis was performed using Matlab programs written in-house.

The LD analysis was performed using Matlab programs written in-house. The model-free fitting was carried out by the Modelfree program (Palmer et al. 1991; Mandel et al. 1995), supplemented by in-house written scripts for

automation and model-selection. To judge the quality of a SDM, LD or MF fit, Monte–Carlo simulations in which the model is fitted to 500 simulated data sets were performed. The Monte–Carlo ensembles were generated so that $y_{\mathrm{fit}} = \langle y \rangle_{\mathrm{MC}}$ and $\sigma_{\mathrm{exp}} = \sigma_{\mathrm{MC}}$. The experimental $\chi^2_{\mathrm{exp}} = \sum_{\mathrm{j}} \left(y_{\mathrm{j,exp}} - y_{\mathrm{j,fit}}\right)^2 / \sigma^2$, where the sum runs over all relaxation rates and all fields for each residue, is then compared to the simulated χ^2_{sim} distribution (Press et al. 1992), If χ^2_{exp} is less than some specified fraction of χ^2_{sim}, the fit is deemed satisfactory. Here, we were somewhat generous in demanding χ^2_{exp} to be less than all of the simulated χ^2_{sim} values (often 95% is used). In order to deduce whether the reduction seen in χ^2 by the inclusion of an extra parameter is justified or not, the F statistic, $F = \upsilon_2/(\upsilon_1 - \upsilon_2) \cdot \left(\chi^2_1 - \chi^2_2\right)/\chi^2_2 = \upsilon_2\left(\chi^2_1 - \chi^2_2\right)/\chi^2_2$, was used (Bevington and Robinson 1992). υ_1 and υ_2 are the number of degrees of freedom for the two models. The confidence level for the inclusion of the extra parameter was then calculated analytically from the theoretical probability function of F, given by (Bevington and Robinson 1992; Press et al. 1992)

$$Q(F|\upsilon_1 - \upsilon_2, \upsilon_2) = I_{\frac{\upsilon_2}{\upsilon_2 + (\upsilon_1 - \upsilon_2)F}}\left(\frac{\upsilon_2}{2}, \frac{\upsilon_1 - \upsilon_2}{2}\right) \quad (16)$$

where $I_x(a,b)$ is the incomplete gamma function. $P = 1 - Q$ is the confidence level for the inclusion of an extra parameter. The theoretical distribution was used with a confidence level of $P = 0.8$.

Results

A full set of ^{15}N laboratory frame relaxation rates (R_1, R_2 and heteronuclear *NOE*) was obtained as described in Material and Methods. The data are shown in Fig. 2. Three features are recognized without considering the absolute values. First, the NOEs are small and nearly always negative at 500 MHz, and increase to be nearly always positive at 800 MHz. Thus, the peptide backbone appears to be very flexible with residue-specific correlation times well below the value of 6.0 ± 0.1 ns for native ACBP (Hansen 2002). In the case of small τ_e, the zero-crossing from negative to positive values in the NOE occurs for a τ_{loc} of about 1 ns. At higher fields, the NOEs become more positive. A large τ_e shifts the NOE to a lower value. Second, both the R_1 and R_2 data show maxima in the regions of the four helices in the native structure, showing that the spectral density at zero and ^{15}N frequency also have maxima there. This is somewhat unusual, since several other denatured proteins show quite flat R_1 profiles, in spite of the presence of peaks in the R_2 data (Alexandrescu and Shortle 1994; Logan et al. 1994; Schwalbe et al. 1997; Schwarzinger et al. 2002). In at least two other studies (Buevich et al. 2001; Ochsenbein et al. 2001), however, the R_1 profile is not flat. In the first,

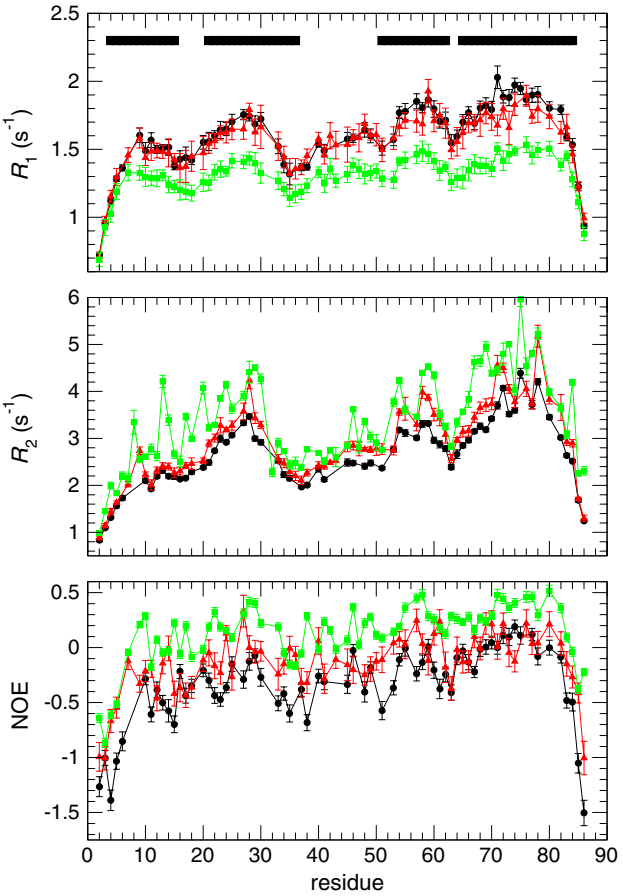

Fig. 2 ^{15}N relaxation data for acid-denatured ACBP acquired at 500 MHz (black circles), 600 MHz (red triangles) and 800 MHz (green squares). The data were acquired at pH 2.3, 27°C. The datasets are, from top to bottom, R_1, R_2 and ^{1}H–^{15}N NOE. The black bars in the topmost panel indicate the location of the α-helices in the native protein

the exchange terms were generally quite large, resulting in higher R_2 rates than in the present study. Correlated peaks in the R_1 and R_2 profiles implicate that τ_{loc} for the "baseline" in the profiles is shorter than 3 ns, since R_1 as a function of τ_{loc} reaches its maximum there (at 500 MHz proton frequency). At longer baseline τ_{loc}, the profiles are anti-correlated, since R_1 gently falls off when τ_{loc} exceeds 3 ns. Third, the scattered peaks in the 600 MHz and 800 MHz R_2 data as compared to the smoother 500 MHz R_2 data indicate that some residues are involved in chemical exchange processes. Since there is a field-dependence in the exchange contribution, the exchange is intermediate to fast on the NMR shift scale.

If τ_e is small, then τ_{loc} can be estimated from the R_2/R_1-ratios (Kay et al. 1989), which are independent of S^2. The estimated τ_{loc} sequence profile for acid-denatured ACBP is shown in Fig. 3. If there are no chemical exchange contributions to R_2, the correlation times should be equal at all fields. The maxima in the correlation times are now most

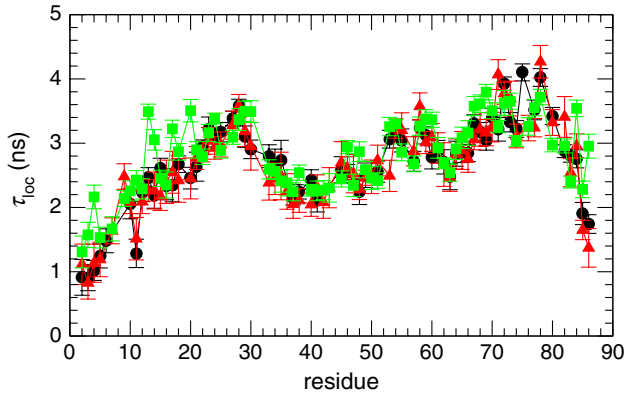

Fig. 3 τ_{loc} for unfolded ACBP, estimated from the R_2/R_1 ratios at 500 MHz (black circles), 600 MHz (red triangles) and 800 MHz (green squares)

clearly pronounced in the regions of helix A2 and A4. A small difference between the three estimated correlation times for a certain residue implies that there is only a small exchange contribution to R_2 for that residue.

Spectral density mapping

As an initial analysis, a reduced spectral density mapping (SDM) was performed as described in Materials and Methods and Appendix. The spectral densities and R_{ex} values determined this way are shown in Fig. 4. The $J(0)$ sequence profile shows the same over-all trends as the estimated τ_{loc}. This is not surprising, since, in the approximation represented by Eq. 4, $J(0) \sim \tau_{loc}$. The trends prevail also for $J(\omega_N)$ sequence profile, which resembles the R_1 profile very much. As discussed above, this is consistent with a quite small "baseline" τ_{loc}, in the range 1–2 ns, and shows that for ACBP the longitudinal relaxation rates describes the overall trends as well as the transverse relaxation rates. The $J(0.87\omega_H)$ profiles are more random, which is also expected since, in the model-free representation, they contain mainly the second term in Eq. 4. The fast motions are most likely not correlated from residue to residue. These profiles are almost identical to the profiles of the cross-relaxation rate σ (not shown). Overall, there is a slight decrease in the $J(0.87\omega_H)$ sequence profile along the protein chain towards the C-terminus. This is consistent with the increases in $J(0)$ and $J(\omega_N)$ profiles observed in the same direction, because, also in the model-free representation, the first term in Eq. 4 falls off more rapidly at high frequencies when τ_{loc} is large. There are significant R_{ex} contributions at many positions along the protein chain. These are not clearly concentrated to any specific region, but there is some weight towards the loop regions between the helices in the native structure.

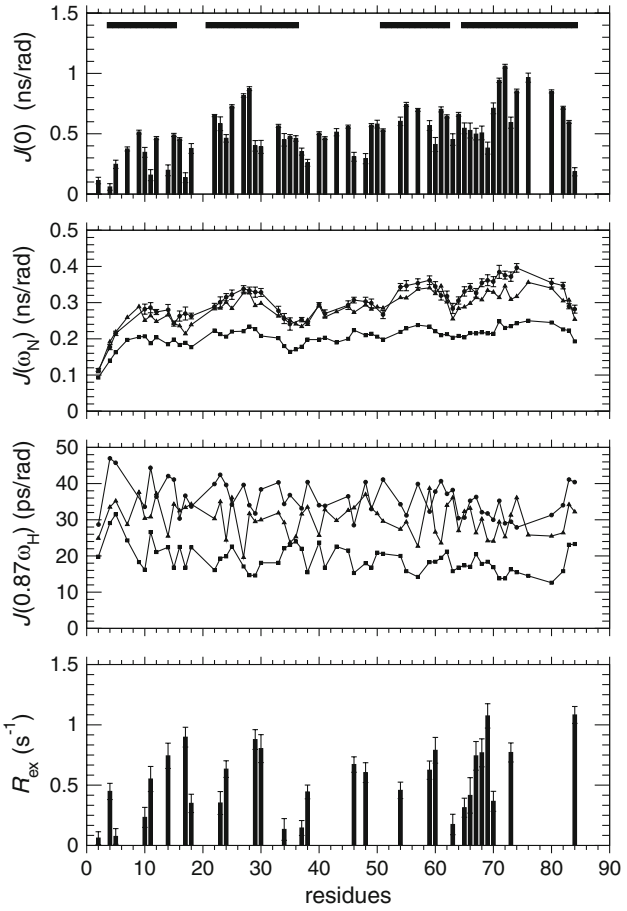

Fig. 4 Reduced spectral density mapping of unfolded ACBP. All the spectral densities and R_{ex} were extracted by linear fitting of all the relaxation data simultaneously as described in text. From top to bottom: $J(0)$; $J(\omega_N)$ at 500 MHz (●), 600 MHz (▲) and 800 MHz (■) proton frequency; $J(0.87\omega_H)$ at 500 MHz (●), 600 MHz (▲) and 800 MHz (■); R_{ex} values corresponding to 500 MHz. To enhance visibility, the standard deviations for $J(\omega_N)$ are drawn only for the 500 MHz data. The black bars in the topmost panel indicate the location of the α-helices in the native protein

Local model-free and LD analyses

A traditional model-free fit with a local correlation time τ_{loc} was performed, following the protocol established by Mandel et al. (1995), with the modification that only the two models involving (τ_{loc}, S^2, $\tau_e > 0$) or (τ_{loc}, S^2, $\tau_e > 0$, R_{ex}) were considered. The reasons for this choice are that the (τ_{loc}, S^2, $\tau_e = 0$) model failed the χ^2 criterion for all residues and that the extended model-free sdf with two order parameters met the χ^2 criterion for only two additional residues (Gln-33 and Gly-63) as compared to the model with R_{ex}. In accordance with Mandel et al., residues that fulfilled the χ^2 criterion for the simpler model were not tested for the higher model. The resulting parameters are shown in Fig. 5. Most of the residues have a τ_e-value of about 100 ps. In the four cases where τ_e is closer to zero

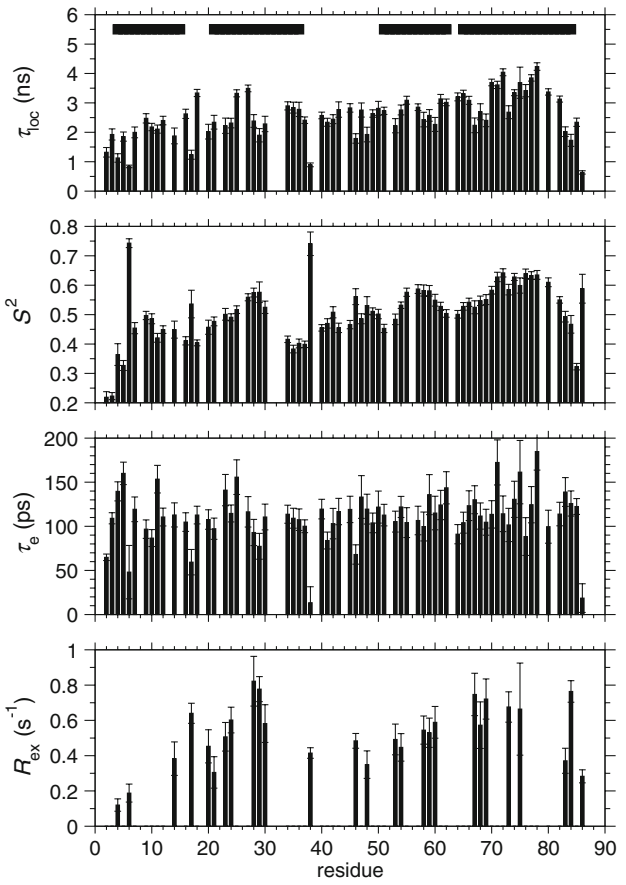

Fig. 5 Model-free relaxation analysis of unfolded ACBP at pH 2.3. The model selection (inclusion of R_{ex} or not) was performed as in Materials and Methods. The R_{ex} values correspond to 500 MHz and the black bars in the topmost panel indicate the location of the α-helices in the native protein

(Asp-6, Thr-17, Asp-38 and Ile-86), S^2 is significantly larger and τ_{loc} smaller than for the rest of the protein. The most apparent of these are Asp-6 and Asp-38. There was significant spectral overlap between Asp-6 and Asp-75 at some of the fields. Hence, fewer relaxation rates could be measured for these two residues. In particular, the NOE at the highest two fields could not be obtained. Since the high-frequency NOEs are needed to properly define the fast correlation time, this could explain the apparently higher S^2 value for Asp-6. In the case of Asp-75, we note that the 600 MHz R_1 value is larger than the value at 500 MHz. This is unphysical and must be due to some unknown systematic error; perhaps overlap with the nearby Asp-48. Apart from these irregularities, both S^2 and τ_{loc} vary rather smoothly along the peptide chain. The local maxima in the R_1 and R_2 relaxation data at the native helix positions are mainly translated into local maxima in the S^2 profile and τ_{loc}, which show increased values towards the C-terminus of the protein. The main reason for performing the model-free analysis is to provide input parameters for the MIC analysis presented below. In order to justify that the model selection was adequately performed, the spectral densities were back-calculated and compared to those determined in the SDM. The spectral density profiles agreed very well, with the same peaks and amplitudes in $J(0)$ and $J(\omega_N)$ (data not shown).

In order to compare the MIC analysis to a model with an analytical distribution function $f(\tau_C)$, we performed a Lorentzian correlation time distribution (LD) analysis (Ochsenbein et al. 2002). The result is displayed in Fig. 6. We chose to limit the fit to the case where $\tau_e = 0$ in Eq. 8, as was done in the original paper. Here, in contrast to that work, the fit is considerably worse than for the local MF fit: only 53 residues meet the χ^2 criteria, as compared to 69 residues in case of the local MF fit. In comparison to the analysis performed in the study by Ochsenbein et al., the local model-free parameters calculated here fall in approximately the same range, except for τ_e, which is slightly larger on average in the present study. Thus it might be that the LD model also needs τ_e, to be non-zero

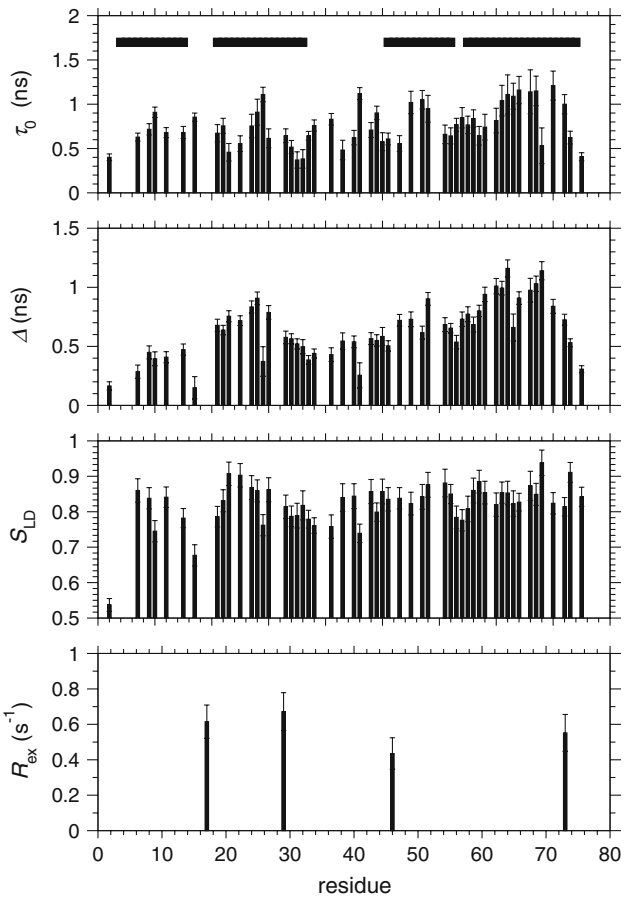

Fig. 6 LD analysis of relaxation data for unfolded ACBP, pH 2.3. The model selection (inclusion of R_{ex} or not) was performed as described in Materials and methods. The R_{ex} values correspond to 500 MHz and the black bars in the topmost panel indicate the location of the α-helices in the native protein

in order to compensate for the large positive skewness of the Lorentzian distribution. The Δ parameter is clearly peaked in the regions of helices A2 and A4, and the order parameter S_{LD} is always well below 1, indicating that the skewness of the fitted distribution $f_{LD}(\tau_C)$, is too positive (see Theory).

MIC analysis

As described in the Theory section, the local model-free parameters shown in Fig. 5 were translated into the MIC parameters $\langle\tau_C\rangle$,σ_τ, γ_τ, $\tau_{1/2}$ and λ (Fig. 7). To estimate the errors, the translation was repeated 500 times with different input values, generated by adding random noise with the same standard deviation as the local MF parameters. As seen, the σ_τ profile is very similar to the τ_{loc} profile of Fig. 5, since S^2 is distributed around $S^2 = 0.5$, so that the larger variations in τ_{loc} dominate (Table 1). The $\langle\tau_C\rangle$ profile is considerably smoother and bears more of the S^2 characteristics. This is in accordance with Table 1: $\langle\tau_C\rangle$ is dominated by the product $\tau_{loc}S^2$, if τ_e is small. Surprisingly, the skewness profile is rather smooth, although this parameter is expected to be determined least accurately. From the theoretical simulations of the skewness it can be expected to be underestimated by approximately 0.5 on average. Thus, the $f(\tau_C)$ distribution is probably most symmetric in the helix A3 and A4 regions and skewed towards short correlation times at the N-terminus (the long-time τ_C tail dominates over the short-time tail), helix A1 and the segment between A2 and A3. The region of helix A2 shows a quite abrupt change in skewness somewhere around residue 32. The same feature is seen in the S^2 profile of Fig. 5. The half-time of the TCF appears to be the most sensitive probe of the dynamical differences over the sequence. The helix A4 region has twice as large half-times as the rest of the protein, but there are also local maxima centered in the first part of helix A2 and the middle of helix A3, respectively. As expected, the TCF integral ratio λ is anti-correlated to $\tau_{1/2}$, which is most clearly seen at the positions of helices A3 and A4. With a few exceptions, λ is in the range 0.85 to 0.9 up to the helix A3 region and drops to below 0.8 at the centre of helix A4.

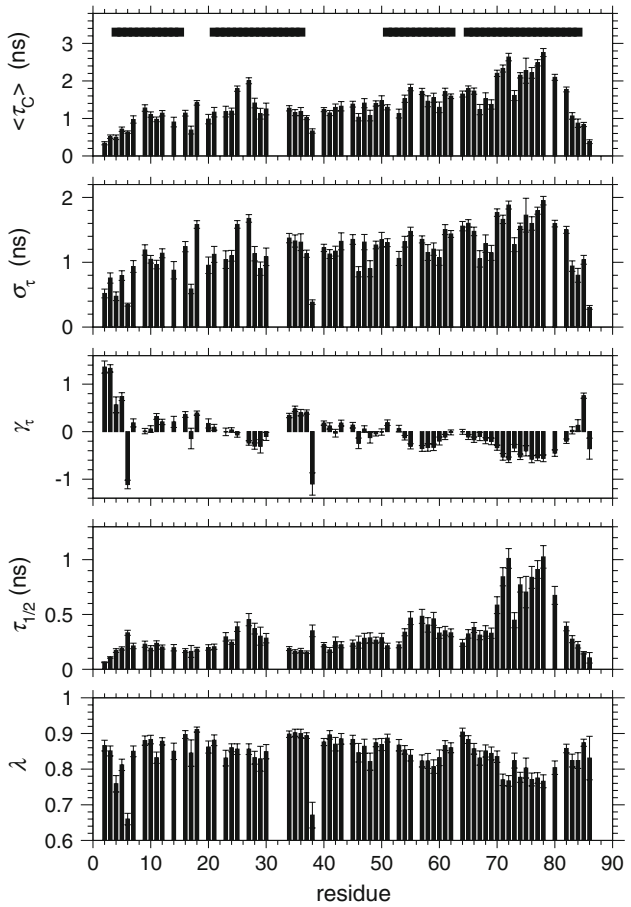

Fig. 7 MIC analysis of relaxation data for unfolded ACBP, pH 2.3. The parameters were calculated from the local model-free parameters in Fig. 5 as described in text. The error bars correspond to one standard deviation and were propagated from the MF parameters using Monte–Carlo simulations. The figure shows from top to bottom: the mean correlation time $\langle\tau_C\rangle$; the standard deviation σ_τ; the skewness γ_τ of the distribution; the half-time $\tau_{1/2}$ of the TCF and the ratio λ between the integral of TCF at $\tau > \tau_{1/2}$ and the integral of the full TCF

Discussion and conclusions

Comparison of model-free, MIC and analytical distribution models

Since an analytical model of $f(\tau_C)$ has a certain shape, its properties cannot be expected to match the MIC parameters perfectly. For example, the LD model has always a large positive skewness when τ_0 is much closer to 0 than to τ_{max}. Still, one would expect some degree of correlation between the corresponding parameters. We show such correlations in Fig. 8, where we compare Δ with σ_τ^{MIC} and $\langle\tau_C\rangle$ and skewness obtained with the two methods. In addition, we compare Δ with σ_τ^{LD}. The latter quantity is defined in the figure caption. As expected, the correlations are best for the mean and width parameters, but very bad for the skewness, which is also greatly over-estimated by the LD analysis. The reason is the long-time tail of $f_{LD}(\tau_C)$. For larger τ_0, the skewness of $f_{LD}(\tau_C)$ gets smaller (not shown). This happens because there is more room for the short-time tail as the maximum of the distribution shifts towards longer values, and does not reflect a real difference in distribution shape.

One advantage of the MIC sculpting of the spectral density is that it can be used to clarify the relationship between the model-free parameters and the shape

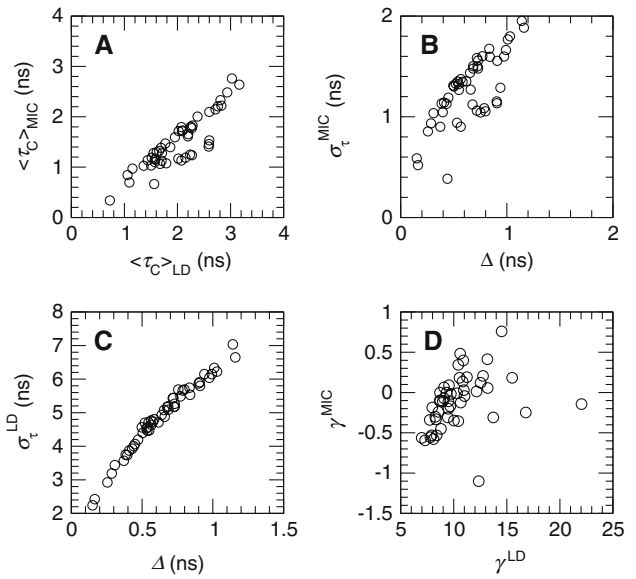

Fig. 8 Correlations between parameters from the MIC analysis and the LD analysis for unfolded ACBP, pH 2.3. The mean and standard deviation, $\sigma_\tau^{\mathrm{LD}} = \left(\langle \tau_C^2 \rangle_{\mathrm{LD}} - \langle \tau_C \rangle_{\mathrm{LD}}^2\right)^{1/2}$, of the Lorentzian distribution were calculated using the relation $\langle \tau_C \rangle_{\mathrm{LD}}^n = \int_0^{\tau_{\max}} f_{\mathrm{LD}}(\tau_C) \tau_C^n \, \mathrm{d}\tau$. (**a**) the correlation between the two mean values; (**b**) the correlation between Δ and $\sigma_\tau^{\mathrm{MIC}}$; (**c**) the correlation between Δ and $\sigma_\tau^{\mathrm{LD}}$; (**d**) the correlation between the two skewness values

parameters of the distribution $f(\tau_C)$. The basis of this was outlined in the Theory section. It was noted in a previous investigation (Ochsenbein et al. 2002), where the partially folded domain 2 of annexin I was studied, that the sequence profile for the width parameter Δ in the LD approach closely followed S_s^2 from an extended model-free fit with a global correlation time τ_m. In that work, τ_{m} was fixed globally to 5 ns, considerably longer than the longest τ_{loc} of 3 ns obtained from a local MF analysis. The order parameter product $S_f^2 S_s^2$ was always well below 0.5 and, since the S_f^2 profile was rather flat, the width of the distribution can therefore be expected to be correlated to $\sqrt{S_s^2}$ (see Table 1). It was also seen that the distribution center τ_0 and the correlation time τ_{s} was correlated, which is also expected from Table 1 if the product $S_f^2 S_s^2$ is small compared to S_f^2, which is the case in the study. Although not obvious from a visual inspection of the data, similar connections can be established between the local MF and LD parameters.

Interpretational ambiguity

The origins of the τ_{C} distribution properties cannot clearly be established at this point. As briefly discussed in the Theory section, there are two different mechanisms that provide a distribution of the correlation times. The first is a spatial separation: different members of the ensemble have different conformations and thereby different local dynamics. The second is a motional mode separation (Ribeiro et al. 1980; Prompers and Brüschweiler 2002): segments of the peptide chain move in correlated fashion due to mutual interactions. Each mode has a certain correlation time and depending on to what degree a NH vector feels each mode, the mode will be weighted differently for each vector. However, there is also the possibility of anisotropic tumbling of each peptide segment. Thus, the skewness can be an indicator of dynamical heterogeneity among the modes or anisotropic segment reorientation. In principle, the modes of motions can be obtained from molecular dynamics simulations and their correlation times subsequently fitted to the relaxation data (Prompers and Brüschweiler 2002). Whether the first or second descriptions apply is mainly dependent on the rate of residual secondary structure formation and disruption. If it is fast enough, the formation itself (which causes fluctuation of the NH vectors) becomes a relaxation mechanism and the mode view is the most appropriate. On the other hand, if the disruption rates are small, a formed cluster will prevail long enough to relax the correlation function as a unit, and the spatial view applies. Some data on the rate on contact formation between residues in a peptide chain (Krieger et al. 2003) and helix unwinding (Lundstrom et al. 2005) exist. In the first study, triplet-triplet energy transfer was used to measure the rate of contact formation between residues separated by peptide chains of different lengths. Over short distances typical time constants of contact were found in the range of 5–10 ns. If one assumes the residual secondary structure to be populated to 20%, this implicates a lifetime for the structure of about 2–3 ns. This can be seen as the diffusion controlled lower bound. In the second study, the unwinding of a helix within a folded protein was studied by $^{13}\mathrm{C}_\alpha(i-1) - {}^{13}\mathrm{C}_\alpha(i)$ differential multiple-quantum spin relaxation. The unwinding was found to occur with a time constant about 40 μs. This is a higher bound of the helix life-time, but the break-up of a transiently formed helical segment can be expected to occur significantly faster. Therefore, with a possible time-span of 2 ns to several μs, we cannot make a clear-cut distinction between the spatial and mode views.

Implications for the nature of the denatured state of ACBP

From secondary chemical shifts and residual dipolar couplings the residual helical content in acid-denatured ACBP at pH 2.3 has been estimated to about 15% (Thomsen et al. 2002; Fieber et al. 2004). Especially, the chemical shift, which is a very sensitive probe of local backbone conformation (Wishart et al. 1992; Wishart and Sykes 1994a, b; Wishart et al. 1995; Schwarzinger et al. 2000;

Schwarzinger et al. 2001), show clear peaks at the helix positions. To elucidate this further we have re-referenced the secondary shifts to the ACBP shifts values obtained at high urea concentrations (Modig et al. 2007). This procedure makes the peaks in the shift patterns even more obvious. Since the amide bond vectors in a residual helix (or a fragment thereof) is expected to reorient more slowly than in fully flexible chain, the residual structure should be evident also in the relaxation data. Indeed, the raw relaxation data (Fig. 2), spectral density mapping (Fig. 4) and local model-free analysis (Fig. 5) display similar peak patterns as the chemical shift data: there are clear peaks at the helix A2, A3 and A4 positions in the R_1, R_2, $J(0)$, $J(\omega_N)$ and S^2 profiles. The MIC analysis shows that these peaks are mainly due to differences in shape of the underlying τ_C distribution (Fig. 7). For example, in helices A3 and A4, the distribution appears skewed towards longer τ_C (negative or zero skewness) on average. This indicates that a significant part of the ensemble is involved in some sort of clustering, for example helix formation, which would result in longer correlation times and possibly also a change in the anisotropy of the segment reorientation. This is further underlined by the simultaneous increase of the half-time, which is very evident in these two regions. In both the γ_τ and $\tau_{1/2}$ profiles, the peaks are centered at the middle of the A3 and A4 helix positions. Interestingly, this is not the case for helix A2, where only the first half show negative skewness and larger than average half-time. Even more intriguing, both peaks are centered at residues 27 or 28. Ile-27 has been showed to form long-range interactions with helix A4 in PRE and RDC studies (Fieber et al. 2004; Kristjansdottir et al. 2005). Thus, the effects seen here strengthen the belief that Ile-27 is important during the initial stages of the folding process. Correlations of urea referenced secondary C^α chemical shifts (Modig et al. 2007) and NH residual dipolar couplings (Fieber et al. 2004) versus skewness and half-time are shown in Fig. 9, together with the linear correlation coefficients. One cannot expect that dynamical and structural properties should correlate perfectly, and hence the correlations are not excellent, but it is surprisingly high for $\tau_{1/2}$ versus $\Delta\delta_{C\alpha}$, and it increases significantly for the skewness when only the helical regions are included. The secondary C^α chemical shifts shown in Fig. 9 were used to derive estimates of the residue specific free energies $\Delta G^\circ = -RT\ln(p_{helix}/p_{rc})$ of transient helix formation. The free energies ranged between 2 and 5 kJ/mol in the natively helical regions, with the lowest values at the positions of helix A4 (Modig et al. 2007). Again, the correlation coefficient between ΔG° and the skewness is 0.65, a further indication that transiently forming structure induces the observed relaxation profiles.

The information content of residual dipolar couplings for denatured proteins are currently under investigation by

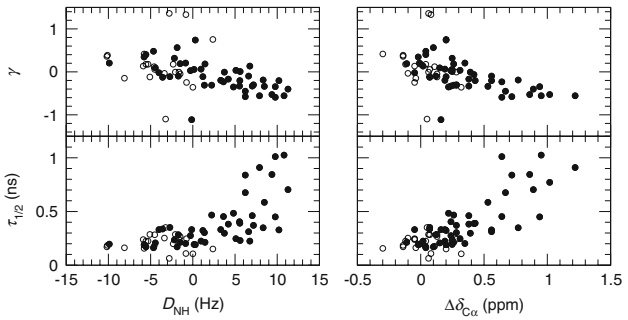

Fig. 9 Correlations of the skewness (top panels) and half-time (lower panels) obtained with MIC analysis of the acid-unfolded state of ACBP versus NH dipolar couplings (Fieber et al. 2004) (left panels) and urea referenced C^α secondary chemical shifts (right panels; (Modig et al. 2007)). Filled circles represent the residues that are helical in the native protein. The correlation coefficients are (natively helical residues within parenthesis): γ vs. D_{NH} −0.47 (−0.61), γ vs. $\Delta\delta_{C\alpha}$ −0.50 (−0.60), $\tau_{1/2}$ vs. D_{NH} 0.68 (0.67) and $\tau_{1/2}$ vs. $\Delta\delta_{C\alpha}$ 0.79 (0.79)

several research groups (Shortle and Ackerman 2001; Louhivouri et al. 2004; Bernado et al. 2005). The current view is that peaks in a residual dipolar coupling pattern may arise for a fully unfolded protein simply because some regions of the dihedral angle space are favored over others (Bernado et al. 2005). This introduces ordered elements that can partly align in the anisotropic medium and thereby introduce peaks in the RDC pattern. To investigate whether this could be the case also for acid denatured ACBP, we calculated the linear correlation coefficient between the NH RDCs and the urea referenced C^α secondary chemical shifts (Modig et al. 2007). The correlation is quite high, $r = 0.82$, thus strongly suggesting that the measured couplings are a consequence of residual secondary structure. The fact that there is a correlation between the D_{NH} data and relaxation data then suggests that both contain partly the same information: a transiently formed structure, generating a peak in the RDC data, would also make the dynamics of the ensemble slower on average. This is not as obvious as it would seem, because the time-scales are different. In order for the transiently formed structures to give measurable RDCs, they must prevail long enough to partly align. Thus, the correlation suggests that at least part of the peaks in the relaxation parameters stem from quite slowly disrupting structures.

It is intriguing to speculate whether transient secondary structure and long-range cluster formation are interdependent or not. The correlations between the relaxation parameters, chemical shifts and RDC values indicate that this is the case. In addition, it has previously been shown that the I27A mutant both decreases the helical content as measured by CD spectroscopy and the long-range effects between the mutated residue and the residues in helix A4 as measured by RDCs (Fieber et al. 2004). However, further

experiments involving other mutations are needed to clarify this issue.

Concluding remarks

In summary, we demonstrate an alternative way of analyzing laboratory frame NMR relaxation data for flexible protein states that only assumes values for the rigid-lattice parameters. The method extracts fundamental statistical properties that are intuitively appealing descriptors of the dynamics within heterogeneous ensembles. Specifically, it is possible to obtain the mean, standard deviation and skewness of the correlation time distribution. In addition, the method has some distinct advantages over other methods that have been used in the analysis of NMR relaxation data of flexible peptide chains. First, the method does not rest on the assumption of time-scale or statistical separation between overall and local motion. Second, it completely avoids the need to specify a functional form of the correlation time distribution. Such forms can be adopted subsequently, if needed for further modeling. Third, the method explains the correlations that are seen between the properties of analytical distributions and the parameters extracted with the traditional model-free approach. Fourth, the statistical quantities can be compared directly to results from molecular dynamics simulations aimed at characterizing how the dynamic modes contribute to the relaxation of each amino-acid residue.

Acknowledgments K. M. is the recipient of a postdoc grant from the Swedish research council. We also thank the John and Birthe Meyer Foundation for the financial support to the structural biology and NMR laboratory (to F. M. P.) and acknowledge the support of the Carlsberg Foundation (to F. M. P.). Dr. Patrik Lundström assisted with the setup of the NMR relaxation experiments.

Appendix

Here we describe a slightly modified way of performing a reduced spectral density mapping (Farrow et al. 1995), which is useful when relaxation data from more than one magnetic field strengths are available. The largest problem with the spectral density mapping is the disentangling of the $J(0)$ and the R_{ex} contributions to R_2. This is often done by a linear fit of R_1, R_2 and $J(0.87\omega_H)$ against the square of the magnetic field, assuming that the system is in fast exchange (Phan et al. 1996). This method is dependent on the assumption that $J(0.87\omega_H)$ can be accurately determined from the dipolar cross-relaxation rate. Instead, we propose to determine the spectral densities $J(0)$, $J(\omega_N^{500})$, $J(\omega_N^{600})$, $J(\omega_N^{800})$, $J(0.87\omega_H^{500})$, $J(0.87\omega_H^{600})$, $J(0.87\omega_H^{800})$ together with R_{ex} at 500 MHz by solving the linear equation system

$$\begin{aligned} R_1^j &= a_{11}J(\omega_N^j) + a_{12}J(0.87\omega_H^j) + a_{13}J(0.87\omega_H^k) \\ R_2^j &= a_{20}J(0) + a_{21}J(\omega_N^j) + a_{22}J(0.87\omega_H^j) \\ &\quad + a_{23}J(0.87\omega_H^k) + a_{24}R_{ex}^{500} \\ \sigma^j &= a_{32}J(0.87\omega_H^j) \end{aligned} \tag{17}$$

with the coefficients

$$\begin{aligned} &a_{11} = \frac{3d^2}{4} + c^2, \quad a_{12} = \frac{7d^2}{4}(1+b_1), \quad a_{13} = -\frac{7d^2}{4}b_1 \\ &a_{20} = \frac{d^2}{2} + \frac{2c^2}{3}, \quad a_{21} = \frac{3d^2}{8} + \frac{c^2}{2}, \quad a_{22} = \frac{13d^2}{8}(1+b_2), \\ &a_{23} = -\frac{13d^2}{8}b_2, \quad a_{24} = \left(\frac{\omega^j}{\omega^{500}}\right)^2 b_2, \quad a_{32} = \frac{5d^2}{4} \\ &b_1 = \frac{(0.921-0.87)\omega_H^j}{0.87(\omega_H^j - \omega_H^k)}, \quad b_2 = \frac{(0.955-0.87)\omega_H^j}{0.87(\omega_H^j - \omega_H^k)} \end{aligned} \tag{18}$$

where the static magnetic field strengths are denoted by j (j = 500, 600 and 800), $d^2 = (\mu_0 h\gamma_N\gamma_H/8\pi^2)^2\langle r^{-3}\rangle^2$, $c^2 = (\Delta\sigma\omega_N)^2/3$ and the index k is explained below. For simplicity, the *NOE* is replaced by the heteronuclear cross-relaxation rate $\sigma = (\gamma_N/\gamma_H)(NOE-1)R_1$. There is one set of such equations for each field j. In Eqs. 17 and 18, the spectral densities $J(\varepsilon\omega_H)$ have been replaced by the leading term in the Taylor expansion of $J(\omega)$. This is the "method 3" described by Farrow et al. (1995), in which the field-dependence of $J(0.87\omega_H)$ is used to determine $J(\varepsilon\omega_H)$:

$$\begin{aligned} J(\varepsilon\omega_H^j) &\approx J(0.87\omega_H^j) + (\varepsilon - 0.87)\omega_H^k \frac{dJ(0.87\omega_H^j)}{d\omega_H^j} \\ \frac{dJ(0.87\omega_H^j)}{d\omega_H^j} &\approx \frac{J(0.87\omega_H^j) - J(0.87\omega_H^k)}{0.87(\omega_H^j - \omega_H^k)} \end{aligned} \tag{19}$$

This is why the index k appears and refers to one of the other fields. To make this approximation as good as possible, the fields j and k were always chosen to be as close as possible, i.e. the pairs 500/600 MHz and 600/800 MHz were used.

The most common way to solve for $J(0)$ and R_{ex}^{500}, which are common to all fields (as seen in Eq. 17, the exchange contribution at the other fields are assumed to scale with the square of the magnetic field strength), is to first determine $J(0.87\omega_H^j)$ from σ^j and then plot a linear combination of R_1^j, R_2^j and $J(0.87\omega_H^j)$ against $(\omega_H^j)^2$ (Phan et al. 1996). $J(\varepsilon\omega_H)$ are estimated by separately using for example Eq. 17 and then inserting the result in the expressions for R_1 and R_2. This procedure has the disadvantage that errors in the NOE data contribute unnecessarily much to the final $J(0)$ and $J(\omega_N)$. Instead, we solve all the equations simultaneously using standard linear algebra. The equation system can be represented as $\mathbf{AJ} = \mathbf{R}$, where $\mathbf{J}$ is a 8 × 1 column vector containing the spectral densities and R_{ex}^{500}, $\mathbf{A}$ is the 9 × 8-matrix of the numerical coefficients and $\mathbf{R}$ is a 9 × 1 column vector

containing the relaxation rates. The system is then easily solved as $\mathbf{J} = \left(\mathbf{A}^{\mathbf{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{R}$. This way, all the relaxation data is used in the determination of all the parameters and the accuracy in $J(0)$ and $J(\omega_N)$ are optimized.

References

Abragam A (1961) Principles of nuclear magnetism. Clarendon Press, Oxford

Ackerman MS, Shortle D (2002) Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence. Biochemistry 41:13791–13797

Alexandrescu AT, Kammerer RA (2003) Structure and disorder in the ribonuclease S-peptide probed by NMR residual dipolar couplings. Protein Sci 12:2132–2140

Alexandrescu AT, Shortle D (1994) Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease. J Mol Biol 242:527–546

Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. Proc Natl Acad Sci 102:17002–17007

Bevington P, Robinson DK (1992) Data reduction and error analysis for the physical sciences. McGraw-Hill Education, NY

Buevich AV, 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 AV, Shinde UP, Inouye M, Baum J (2001) Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. J Biomol NMR 20:233–249

Cao W, Bracken C, Kallenbach NR, Lu M (2004) Helix formation and the unfolded state of a 52-residue helical protein. Protein Sci 13:177–189

Choy WY, Kay LE (2003) Probing residual interactions in unfolded protein states using NMR spin relaxation techniques: an application to Δ131Δ. J Am Chem Soc 125:11988–11992

Choy W-Y, Mulder FAA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay J, Kay LE (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. J Mol Biol 316:101–112

Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. J Am Chem Soc 112:4989–4991

Cole KS, Cole RH (1941) Dispersion and absorption in dielectrics. I. Alternating current characteristics. J Chem Phys 9:341–351

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 33:5984–6003

Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE (1995) Spectral density function mapping using ^{15}N relaxation data exclusively. J Biomol NMR 6:153–162

Favro LD (1960) Theory of the rotational brownian motion of a free rigid body. Phys Rev 119:53–62

Ferrarini A, Nordio PL, Moro GJ (1994) The molecular dynamics of liquid crystals. Kluwer Academic, Dordrecht

Fieber W, Kristjansdottir S, Poulsen FM (2004) Short-range, long-range and transition state interaction in the denatured state of ACBP from residual dipolar couplings. J Mol Biol 339:1191–1199

Fieber W, Kragelund BB, Meldal M, Poulsen FM (2005) Reversible dimerisation of acid-denatured ACBP controlled by helix A4. Biochemistry 44:1375–1384

Goddard TD, Kneller DG (2005) Sparky 3. San Francisco, University of California

Halle B, Wennerström H (1981) Interpretation of magnetic resonance data from water nuclei in heterogenous systems. J Chem Phys 75:1928–1943

Halle B, Johannesson H, Venu K (1998) Model-free analysis of stretched relaxation dispersions. J Magn Reson 135:1–13

Halle B, Denisov VP, Modig K, Davidovic M (2005) Protein conformational transitions as seen from the solvent: magnetic relaxation dispersion studies of water, cosolvent and denaturant interactions with nonnative proteins. In: Kiefhaber T (ed) Protein folding handbook. Wiley, UK

Hansen PI (2002) NMR ^{15}N-relaxation studies of acyl-coenzym A-binding protein at multiple magnetic field strengths. Thesis for M.Sc. Chemical Institute, University of Copenhagen, Copenhagen

Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy. Biochemistry 28:8972–8979

Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. Science 295:1719–1722

Kragelund BB, Robinson CV, Knudsen J, Dobson CM, Poulsen FM (1995) Folding of a four-helix bundle: studies of acyl-coenzyme A binding protein. Biochemistry 34:7117–7224

Kragelund BB, Osmark P, Neergaard TB, Schiødt J, Kristiansen K, Knudsen J, Poulsen FM (1999) The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. Nature Struct Biol 6:594–601

Krieger F, Fierz B, Bieri O, Drewello M, Kiefhaber T (2003) Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. J Mol Biol 332:265–274

Krieger F, Moglich A, Kiefhaber T (2005) Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. J Am Chem Soc 127:3346–3352

Kristjansdottir S, Lindorff-Larsen K, Fieber W, Dobson CM, Vendruscolo M, Poulsen FM (2005) Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies. J Mol Biol 347:1053–1062

Kumar S, Modig K, Halle B (2003) Trifluoroethanol-induced $\beta \rightarrow \alpha$ transition in β-lactoglobulin: hydration and cosolvent binding studied by 2H, 17O, and 19F magnetic relaxation dispersion. Biochemistry 42:13708–13716

Lindorff-Larsen K, Kristjansdottir S, Teilum K, Fieber W, Dobson CM, Poulsen FM, Vendruscolo M (2004) Determination of an ensamble of structures representing the denatured state of the bovine Acyl-Coenzyme A binding protein. J Am Chem Soc 126:3291–3299

Lindorff-Larsen K, Rogen P, Paci E, Vendruscolo M, Dobson CM (2005) Protein folding and the organization of the protein topology universe. Trends Biochem Sci 30:13–19

Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J Am Chem Soc 104:4546–4559

Logan TM, Thériault Y, Fesik SW (1994) Structural chracterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. J Mol Biol 236:637–648

Louhivouri M, Fredriksson K, Pääkkönen K, Permi P, Annila A (2004) Alignment of chain-like molecules. J Biomol NMR 29:517–524

Lundstrom P, Mulder FA, Akke M (2005) Correlated dynamics of consecutive residues reveal transient and cooperative unfolding of secondary structure in proteins. Proc Natl Acad Sci 102: 16984–16989

Mandel AM, Akke M, Palmer AG (1995) Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. J Mol Biol 246:144–163

Mandrup S, Hojrup P, Kristiansen K, Knudsen J (1991) Gene synthesis, expression in *Escherichia coli*, purification and charecterization of the recombinant bovine acyl-CoA-binding protein. Biochem J 276:817–823

Millet IS, Doniach S, Plaxco KW (2002) Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. Adv Protein Chem 62:241–262

Modig K, Kurian E, Prendergast FG, Halle B (2003) Water and urea interactions with the native and unfolded forms of a β-barrel protein. Protein Sci 12:2768–2781

Modig K, Jurgensen VW, Lindorff-Larsen K, Fieber W, Bohr HG, Poulsen FM (2007) Detection of initiation sites in protein folding of the four helix bundle ACBP by chemical shift analysis. FEBS Lett 581:4965–4971

Ochsenbein F, Neumann JM, Guittet E, van Heijenoort C (2002) 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. Protein Sci 11:957–964

Ochsenbein F, Guerois R, Neumann JM, Sanson A, Guittet E, van Heijenoort C (2001) ^{15}N 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

Palmer AGIII (2004) NMR characterization of the dynamics of biomacromolecules. Chem Rev 104:3623–3640

Palmer AG, Rance M, Wright PE (1991) Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance carbon-13 heteronuclear NMR spectroscopy. J Am Chem Soc 113:4371–4380

Phan IQH, Boyd J, Campbell ID (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 WH, Flannery BP, Teukolsky SA, Vetterling WT (1992) Numerical recipes in C: the art of scientific computing. Cambridge University Press, Cambridge

Prompers JJ, Brüschweiler R (2002) General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. J Am Chem Soc 124:4522–4534

Ribeiro AA, King R, Restivo C, Jardetzky O (1980) An approach to the mapping of internal motions in proteins. Analysis of ^{13}C NMR relaxation in the bovine pancratic trypsin inhibitor. J Am Chem Soc 102:4040–4051

Sallum CO, Martel DM, Fournier RS, Matousek WM, Alexandrescu AT (2005) Sensitivity of NMR residual dipolar couplings to perturbations in folded and denatured staphylococcal nuclease. Biochemistry 44:6392–6403

Sanchez IE, Kiefhaber T (2003) Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. J Mol Biol 327: 867–884

Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM (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

Schwarzinger S, Kroon GJ, Foss TR, Wright PE, Dyson HJ (2000) Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. J Biomol NMR 18:43–48

Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ (2001) Sequence-dependent correction of random coil NMR chemical shifts. J Am Chem Soc 123:2970–2978

Schwarzinger S, Wright PE, Dyson HJ (2002) Molecular hinges in protein folding: the urea-denatured state of Apomyoglobin. Biochemistry 41:12681–12686

Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552

Shojania S, O'Neil JD (2006) HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1-72) by NMR spectroscopy. J Biol Chem 281:8347–8356

Shortle D (1996) Structural analysis of non-native states of proteins by NMR methods. Curr Opin Struct Biol 6:24–30

Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. Science 293:487–489

Teilum K, Kragelund BB, Knudsen J, Poulsen FM (2000) Formation of hydrogen bonds precedes the rate-limiting formation of persistent structure in the folding of ACBP. J Mol Biol 301:1307–1314

Teilum K, Kragelund BB, Poulsen FM (2002) Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. J Mol Biol 324:349–357

Teilum K, Poulsen FM, Akke M (2006) The inverted chevron plot measured by NMR relaxation reveals a native-like unfolding intermediate in acyl-CoA binding protein. Proc Natl Acad Sci 103:6877–6882

Thomsen JK, Kragelund BB, Teilum K, Knudsen J, Poulsen FM (2002) Transient intermediary states with high and low folding probabilities in the apparent two-state folding equilibrium of ACBP at low pH. J Mol Biol 318:805–814

van Kampen NG (1981) Stochastic processes in physics and chemistry. North-Holland, Amsterdam

Weisstein EW (2006a) Central moment. In: MathWorld-A Wolfram Web Resource. http://mathworld.wolfram.com/CentralMoment.html Cited 6 February 2008

Weisstein EW (2006b) Skewness. In: MathWorld-A Wolfram Web Resource. http://mathworld.wolfram.com/Skewness.html Cited 6 February 2008

Wishart DS, Sykes BD (1994a) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 4:171–180

Wishart DS, Sykes BD (1994b) Chemical shifts as a tool for structure determination. Methods Enzymol 239:363–392

Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry 31:1647–1651

Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J Biomol NMR 5:67–81

Woessner DE (1962) Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. J Chem Phys 37:647–654

Zhang X, Xu Y, Zhang J, Wu J, Shi Y (2005) Structural and dynamic characterization of the acid-unfolded state of hUBF HMG box 1 provides clues for the early events in protein folding. Biochemistry 44:8117–8125