Journal of Biomolecular NMR, **20:** 233–249, 2001.
KLUWER/ESCOM

Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies

Alexei V. Buevich[a], Ujwal P. Shinde[b], Masayori Inouye[b] & Jean Baum[a,*]
[a]*Department of Chemistry, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854, U.S.A.;* [b]*Department of Biochemistry, Robert Wood Johnson Medical School-UMDNJ, 675 Hoes Lane, Piscataway, NJ 08854, U.S.A.*

Received 24 November 2000; Accepted 23 April 2001

Key words: backbone protein dynamics, Cole–Cole model free analysis, model free analysis, ^{15}N relaxation, natively unfolded proteins, spectral density mapping

Abstract

The dynamics of the natively unfolded form of the pro-peptide of subtilisin (PPS) have been characterized at two different pHs (6.0 and 3.0) by ^{15}N relaxation experiments. ^{15}N relaxation data is obtained at multiple field strengths and a detailed comparison of spectral density mapping, the model free approach and the recently proposed Cole–Cole model free (CC-MF) analysis is presented. The CC-MF analysis provides a better fit to the observed magnetic field dependence of ^{15}N relaxation data of unfolded PPS than conventional model free approaches and shows that fluctuations in R_2 may be accounted for by a distribution of correlation times on the nanosecond timescale. A new parameter ε derives from the analysis and represents the width of the distribution function and the heterogeneity of the dynamics on the nanosecond timescale at a particular site. Particularly interesting is the observation that ε is sensitive to pH changes and that PPS samples a wider distribution of nanosecond time scale motions at less acidic pHs than at more acidic pHs. These results suggest that PPS experiences a higher degree of correlated motion at pH 6.0 and that electrostatic interactions may be important for inducing correlated motions on the nanosecond timescale in unfolded PPS.

Abbreviations: PPS, pro-peptide of subtilisin; KDH, Kyte–Doolittle hydropathy; MF, model free; CC-MF, Cole–Cole model free; HSQC, heteronuclear single-quantum coherence; HMQC, heteronuclear multiple-quantum coherence; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; R_1, longitudinal relaxation rate; R_2, transverse relaxation rate; CPMG, Carr–Purcell–Meiboom–Gill.

Introduction

It has recently been found that a number of gene sequences in the genome code for protein sequences that are likely to be unfolded or intrinsically unstructured (Wright and Dyson, 1999; Uversky et al., 2000). These proteins are involved in important regulatory functions and may only fold into three-dimensional structures upon recognition of their target protein. Understanding the dynamics of these unfolded states can provide valuable insight into protein function (Dill and Shortle, 1991; Shortle, 1996), protein diseases such as amyloidoses and prion disease (Cohen and Prusiner, 1998), and the early stages of protein folding (Dobson et al., 1998; Dyson and Wright, 1998). Unlike the folded state, unfolded proteins cannot be adequately described by a single conformation, but rather by an ensemble of interconverting conformations. The conformational dynamics are an integral component of this ensemble, and are critical for understanding the nature of the unfolded state (Dyson and Wright, 1998).

NMR spectroscopy plays a leading role in the analysis of the dynamics of folded proteins and more recently of unfolded and partially folded proteins. The NMR detectable molecular motions range from picosecond–nanosecond and

*To whom correspondence should be addressed. E-mail: baum@rutchem.rutgers.edu

microsecond–millisecond time scales to very slow conformational interconversions on the time scale of seconds, minutes or even hours (Palmer et al., 1996; Dobson and Hore, 1998; Kay, 1998). In unfolded and partially folded proteins, NMR experiments have allowed definition of residual structure and reduced conformational dynamics (Neri et al., 1992; Alexandrescu and Shortle, 1994; Arcus et al., 1995; Farrow et al., 1995; Frank et al., 1995; Zhang and Forman-Kay, 1995; Buck et al., 1996; Wong et al., 1996; Brutscher et al., 1997; Schwalbe et al., 1997; Eliezer et al., 1998; Penkett et al., 1998; Meekhof and Freund, 1999). However, NMR relaxation data of unfolded proteins are difficult to analyze quantitatively using standard approaches. It has been recognised that the assumptions of the model free analysis of Lipari and Szabo (Lipari and Szabo, 1982a,b) are more suitable for folded proteins and may not be appropriate for unfolded proteins which are composed of an ensemble of states (Frank et al., 1995; Brutscher et al., 1997; Farrow et al., 1997; Penkett et al., 1998). Recently, we have proposed a novel approach to the analysis of NMR relaxation data that incorporates the notion that individual sites in unfolded proteins need to be described by a distribution of correlation times on the nanosecond time scale. This newly developed Cole–Cole model free (CC-MF) analysis establishes the width of the distribution of correlation times and describes the heterogeneity of the dynamics of unfolded states more quantitatively (Buevich and Baum, 1999).

The 77-residue pro-peptide of subtilisin (PPS) has been shown to be unfolded under physiological conditions (Shinde et al., 1993). PPS functions as an intramolecular chaperone that facilitates correct folding of the catalytic domain of subtilisin and also acts as a competitive inhibitor of active subtilisin (Zhu et al., 1989). In vivo, subtilisin exists as a precursor, namely as pre-pro-subtilisin. Upon completion of folding, the precursor removes the intramolecular chaperone domain through autoproteolysis to give active subtilisin (Ikemura et al., 1987). It has been shown that folding of denatured subtilisin in the absence of the pro-peptide traps the protease into a molten globule-like intermediate (Eder et al., 1993). With the *trans* addition of the pro-peptide the intermediate can adopt an active conformation (Shinde and Inouye, 1995). These data suggest that the information required to fold the subtilisin protein is encoded in the amino acid sequence of the pro-peptide alone.

Here we report a detailed study of the dynamics of the natively unfolded form of the pro-peptide of subtilisin and explore the role of electrostatics in defining the dynamical heterogeneity of the unfolded form. A full description of different approaches to the quantitative analysis of unfolded proteins is presented and we show that the newly developed CC-MF approach affords the best fit to the relaxation data obtained at multiple field strengths. We show that the new relaxation parameter ε, which represents the distribution of correlation times on the nanosecond time scale, is sensitive to the ionization state of the protein and allows us to characterize variations in the dynamics of PPS as a function of pH.

Materials and methods

Expression and purification

The 77-mer pro-peptide from subtilisin E was expressed in *Escherichia coli* using a pET11a expression vector under the control of a T7-promoter (Li and Inouye, 1994). When preparing the ^{15}N-labelled peptide, the clone was grown and expressed in minimal media with $^{15}NH_4Cl$ as the only nitrogen source. The inclusion bodies were solubilized in 15–25 ml 6 M guanidine-HCl and after overnight incubation at 4 °C the insoluble material was removed by centrifugation. The supernatant was dialysed against 50 mM sodium phosphate buffer at pH 6.0. The purification steps included a cation-ion-exchange column (CM-Sephadex-200) with a linear gradient of NaCl (0–0.4 M) followed by HPLC (C18-reverse phase) with a linear acetonitrile gradient (0–65%). The pro-peptide was eluted at 0.2 M NaCl and 43% acetonitrile with two purification steps.

NMR experiments

The NMR samples were prepared in aqueous solution (10% D_2O) with 50 mM sodium phosphate buffer at pH 6.0 at a concentration of 0.7 mM.

The NMR measurements were carried out at 9 °C on Varian Inova-600, Inova-500 and Unity-400 spectrometers operating at proton frequencies of 599.926, 499.938 and 399.945 MHz, respectively. The ^{1}H chemical shifts of the water signal were used as a reference: 5.0 ppm at 9 °C (Wishart et al., 1995). The ^{15}N chemical shifts were calculated using a Ξ ratio of $^{15}N/^1H$ equal to 0.101329118 (Wishart et al., 1995).

The two-dimensional (2D) ^{1}H–^{15}N HSQC experiments on the ^{15}N labelled samples were performed by a sensitivity-enhanced gradient selection method (Kay et al., 1992). 128 complex points in t_1 (^{15}N) and

1024 complex points in t_2 (^{1}H) were acquired. The t_1-dimension included linear prediction to 256 complex points. Zero-filling to 512 (in t_1) and 2048 (in t_2) complex points was used. A Kaizer-type window function in both dimensions was applied prior to Fourier transformation.

The three-dimensional (3D) NOESY-HSQC and TOCSY-HSQC experiments were performed in the States mode for the NOESY (300 ms) and TOCSY (80 ms) components and in a sensitivity-enhanced gradient selection mode for the HSQC component. The numbers of acquired complex points were $1024\times96\times32$ in the t_3 (^{1}H), t_2 (^{1}H) and t_1 (^{15}N) dimensions, respectively. Sixteen scans (in NOESY-HSQC) and 8 scans (in TOCSY-HSQC) per increment with 1 s delay between scans were used. Spectral widths were 5000 Hz and 1400 Hz in the ^{1}H and ^{15}N dimensions, respectively. Linear prediction and zero-filling in the indirect dimensions gave a matrix of $1024\times256\times128$ complex points. A Kaizer-type window function in all three dimensions was used prior to Fourier transformation.

The 3D HMQC-NOESY-HSQC spectrum (Frenkiel et al., 1990; Ikura et al., 1990) was acquired in TPPI mode during the HMQC component and in gradient selection mode during the HSQC. The numbers of acquired complex points were $1024\times64\times32$ in the t_3 (^{1}H), t_2 (^{15}N) and t_1 (^{15}N) dimensions, respectively. A 300 ms mixing time, 16 scans per increment and a 1 s delay between scans were used. Linear prediction and zero-filling in the indirect dimensions gave a final matrix of $1024\times256\times128$ complex points. Spectral widths and window functions were the same as in previously described 3D experiments.

^{15}N R_1, R_2 and $\{^1$H$\}$-^{15}N NOE measurements were carried out by standard pulse sequences described elsewhere (Palmer, 1993; Farrow et al., 1994). Ten relaxation delays (0.02, 0.04, 0.08, 0.16, 0.32, 0.48, 0.64, 0.80, 1.00 and 1.40 s) were used to measure R_1 values, and R_2 values were measured with 0.008, 0.016, 0.032, 0.04, 0.056, 0.08, 0.12, 0.2, 0.32 and 0.4 s relaxation delays. Recycle delays were 2.5 s in the R_1, 2 s in the R_2 and 5 s in the $\{^1$H$\}$-^{15}N NOE experiment. In the R_2 measurements, a 1.8 ms delay was used in the CPMG pulse train between successive applications of ^{15}N 180° pulses to minimize sample heating. This relatively long delay could result in additional contributions to the measured R_2 values due to proton–proton relaxation (Peng and Wagner, 1992); however, calculations indicate that this contribution is expected to be less than 2%. The $\{^1$H$\}$-^{15}N NOE measurements were performed by 3 s high power pulse train saturation within a 5 s recycle delay. Sixteen scans in R_1, R_2 and 32 scans in $\{^1$H$\}$-^{15}N NOE spectra per t_1 experiment were acquired. In total 2048×128 complex points were obtained for R_1, R_2 experiments and 1024×100 complex points for $\{^1$H$\}$-^{15}N NOE experiments.

The processing of 2D and 3D spectra was performed by Felix-97 software (Molecular Simulation, Inc.).

^{15}N Relaxation data analysis

^{15}N R_1 and R_2 relaxation rates were estimated by least squares minimization fitting of the measured peak heights to a two-parameter equation:

$$I(t) = I_0 \exp(-tR_{1,2}) \quad (1)$$

where I(t) is the intensity of the peak after the relaxation delay t and I_0 is the peak intensity at zero time. The standard deviations of R_1 and R_2 rates were obtained from the covariance matrix for the least squares optimization of a parameter and were on average around 1 and 5%, respectively. The $\{^1$H$\}$-^{15}N NOE values were calculated as follows:

$$NOE = I_{noe}/I_{st} \quad (2)$$

where I_{noe} is the peak intensity with proton presaturation and I_{st} without. Due to the low intensity of the peak intensities in the experiment with proton presaturation, the accuracy of NOE values was assumed to be 5% . However, high reproducibility of the NOE values measured at three field strengths suggested that the actual accuracy might be higher than the approximated value of 5%. All calculations were performed with the Excel program (Microsoft).

Spectral density mapping and model free analysis.

Reduced spectral density mapping was calculated as described by Farrow et al. (1995) using an in-house Excel program (Microsoft). Model free analysis was performed using the Model Free 3.1 program (Dr. A. Palmer, Columbia University). The accuracy of the relationship between the experimental relaxation data and the theoretical data predicted from the results of model free analysis can be estimated by the R-factor, a dimensionless parameter, used in the analysis of spectroscopic data:

$$R = \sqrt{\sum_i (Y_i^{exp} - Y_i^{teor})^2 / \sum_i (Y_i^{exp})^2}$$

where Y_i^{exp} and Y_i^{teor} are the ith experimental and theoretical relaxation data (R_1, R_2, {^{1}H}-^{15}N NOE), respectively; i is the residue number index ($1 \leq i \leq 55$).

Cole–Cole model free analysis

Recently we proposed a variation of the model free approach that incorporates the notion that unfolded proteins are composed of an ensemble of states and that characterizes the dynamics of individual sites by a distribution of correlation times on the nanosecond timescale (Buevich and Baum, 1999). The parameters S^2, τ_0, ε, τ_e and R_{ex} were optimized separately for each residue by performing a non-linear minimization of the theoretical R_1, R_2 and {^{1}H}-^{15}N NOE values at three fields to the experimental data, Equation 16, using an in-house Excel program (Microsoft). The approach to the analysis of uncertainties of the Cole–Cole parameter estimations was done similarly to that used in the Model Free program (Dr. A. Palmer, Columbia University). The experimental uncertainties used in the error analysis were conservatively chosen to be 5% for R_1, R_2 and NOE. Then a Monte Carlo simulation generated 'experimental' values of R_1, R_2 and NOE by sampling an even distribution within these experimental errors. For each amino acid a set of 90 experimental values of R_1, R_2 and NOE was generated and subsequently analyzed using a minimization algorithm. Then the set of Cole–Cole parameters was analyzed to estimate the standard deviation for each iterated parameter. The reproducibility of the error estimations was checked by repeating the analysis on a second set of 90 trial experimental data. This test was done for several amino acids and due to the high reproducibility of the results, only one set of trial experimental values was used to estimate parameter uncertainties. We also tested our approach against the Model Free program, by performing the model free analyses with zero widths of the Cole–Cole distributions. The error estimations were very similar to or in some instances even larger than those found by the Monte Carlo algorithm used in the Model Free program.

Background theory

The relaxation rates for ^{15}N nuclei are given by (Abragam, 1961):

$$R_1 = (d/2)^2\{J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\} + c^2 J(\omega_N) \quad (3)$$

$$R_2 = 1/2(d/2)^2\{4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\} + (1/6)c^2\{4J(0) + 3J(\omega_N)\} + R_{ex} \quad (4)$$

$$NOE = 1 + (d/2)^2(\gamma_H/\gamma_N)\{6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\}/R_1 \quad (5)$$

where $d = \mu_0 h \gamma_N \gamma_H \langle r_{NH}^{-3} \rangle / (8\pi^2)$, $c = 3^{-1/2}\omega_N \Delta\sigma$, μ_0 is the permeability of free space, h is Planck's constant; γ_N and γ_H are the gyromagnetic ratios of ^{15}N and ^{1}H, respectively; $r_{NH} = 1.02$ Å is the N-H bond length; ω_N and ω_H are the Larmor frequencies of ^{15}N and ^{1}H, respectively; $\Delta\sigma = -160$ ppm (Hiyama et al., 1988). R_{ex} is the contribution of chemical exchange processes on the microsecond–millisecond time scale to the transverse relaxation rate R_2. For exchange between two sites, A and B, R_{ex} is given by (Ishima and Torchia, 1999):

$$R_{ex} = (\Delta\omega)^2 p_A p_B \tau_{ex}\{1 - (\tau_{ex}\omega_e/3^{1/2}) \tanh(\tau_{ex}\omega_e/3^{1/2})\} \quad (6)$$

where p_A and p_B are populations of sites A and B, $\Delta\omega = \omega_A - \omega_B$ is the chemical shift difference between sites A and B, $\tau_{ex} = (1/k_{A\to B} + 1/k_{B\to A})$ is the time constant of the exchange process, and $\omega_e = 3^{1/2}/\tau_{CPMG}$, where τ_{CPMG} is one-half of the time delay between 180° pulses in the CPMG pulse train.

Spectral density mapping has been applied to the analysis of unfolded proteins in order to describe their intramolecular motions without using any assumptions about a specific molecular model (Farrow et al., 1995a,b; Ishima and Nagayama, 1995; Peng and Wagner, 1995). Reduced spectral density mapping for ^{15}N spins provides values of J(0), J(ω_N), and J($0.87\omega_H$) from R_1, R_2 and the ^{1}H–^{15}N cross relaxation rate constant σ equal to $(NOE - 1)R_1\gamma_N/\gamma_H$. The reduced spectral density values are given by:

$$J(0.87\omega_H) = 4\sigma/(5d^2) \quad (7)$$

$$J(\omega_N) = [4R_1 - 5.00\sigma]/[3d^2 + 4c^2] \quad (8)$$

$$J(0) = [6R_2 - 3R_1 - 2.72\sigma]/[3d^2 + 4c^2] \quad (9)$$

Slow micro- to millisecond motions may be reflected in J(0) spectral density values as an increase in J(0) values.

The model free formalism of Lipari and Szabo (1982a, b) has been used to describe the motions of folded proteins for many years and more recently has

been applied to defining the dynamics of unfolded proteins. The spectral density function used in the model free approach assumes a separation between the overall rotational motion and the internal motions:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{\left(1-S^2\right)\tau}{1+(\omega\tau)^2}\right] \quad (10)$$

where $\tau^{-1} = \tau_e^{-1} + \tau_m^{-1}$, S^2 is the generalised order parameter, τ_m is the correlation time for overall tumbling, τ_e is the correlation time for internal motion. Equation 10 can include an additional fast motion with a correlation time substantially shorter than τ_e (Clore et al., 1990):

$$J(\omega) = \frac{2}{5}S_f^2\left[\frac{S_s^2\tau_m}{1+(\omega\tau_m)^2} + \frac{\left(1-S_s^2\right)\tau}{1+(\omega\tau)^2}\right] \quad (11)$$

where $\tau^{-1} = \tau_e^{-1} + \tau_m^{-1}$, S_s^2 and S_f^2 are generalised order parameters of slow and fast motion and $S^2 = S_s^2 S_f^2$ when these two motions are independent and axially symmetric.

Recently we have proposed a variation of the model free approach, the Cole–Cole model free approach that characterizes the dynamics of individual sites by a distribution of correlation times on the nanosecond time scale (Buevich and Baum, 1999). Similar approaches were developed for the analysis of R_1 and NOE data of polymers, including Cole–Cole (Cole and Cole, 1941), Fuoss–Kirkwood (Fuoss and Kirkwood, 1941) and (log-χ^2) (Shaefer, 1973) distributions. We incorporated the simplest symmetrical distribution function, the Cole–Cole distribution function, into the model free approach (termed CC-MF). The Cole–Cole distribution function, F(s), is given by:

$$F(s) = \frac{1}{2\pi}\frac{\sin(\varepsilon\pi)}{\cosh(\varepsilon s) + \cos(\varepsilon\pi)} \quad (12)$$

where $s = \ln(\tau_c/\tau_0)$. τ_0 and ε ($0<\varepsilon<1$) are respectively the mean value and the width of the distribution function (Figure 1a). The width at half height of the Cole–Cole distribution is $\Delta s_{1/2} = (2/\varepsilon)\cosh^{-1}(2 + \cos(\pi\varepsilon))$ (in the time domain $\Delta\tau_{1/2} = (2\tau_0)\sinh(1/\varepsilon)\cosh^{-1}(2+\cos(\pi\varepsilon))$), so a smaller value of ε corresponds to a wider distribution of correlation times and $\varepsilon = 1$ corresponds to the single correlation time model characterised by τ_0. Fourier transformation of the Cole–Cole distribution function gives rise to the spectral density function (Figure 1b):

$$J(\omega) = \frac{1}{\omega}\frac{\cos\left[\frac{\pi}{2}(1-\varepsilon)\right]}{\cosh\left[\varepsilon\ln(\omega\tau_0)\right] + \sin\left[\frac{\pi}{2}(1-\varepsilon)\right]} \quad (13)$$

The spectral density function, based on the Cole–Cole distribution function, was incorporated into the model free approach. The new Cole–Cole model free spectral density function was readily obtained by substituting the two Lorentzian functions in Equation 10 by two Cole–Cole functions (Equation 13) giving:

$$J(\omega) = \frac{1}{5}\left[\frac{1}{\omega}\frac{S^2\cos\left[\frac{\pi}{2}(1-\varepsilon)\right]}{\cosh\left[\varepsilon\ln(\omega\tau_0)\right] + \sin\left[\frac{\pi}{2}(1-\varepsilon)\right]} + \frac{1}{\omega}\frac{(1-S^2)\cos\left[\frac{\pi}{2}(1-\varepsilon_e)\right]}{\cosh\left[\varepsilon_e\ln(\omega\tau)\right] + \sin\left[\frac{\pi}{2}(1-\varepsilon_e)\right]}\right] \quad (14)$$

where $\tau^{-1} = \tau_0^{-1} + \tau_e^{-1}$. Equation 14 has two terms and can be further simplified by assuming $\varepsilon_e = 1$, indicating a distribution of local 'overall' correlation times on the nanosecond time scale. This assumption is supported by the fact that use of the 'extended model free approach' with two distinct correlation times to describe the internal motions did not improve the fit of the relaxation data (Clore et al., 1990) (Equation 11) (see Supplementary material). Thus, while ε was optimised, ε_e remained fixed at 1 resulting in degeneracy of this term into a Lorenzian function with a single correlation time:

$$J(\omega) = \frac{1}{5}\left[\frac{1}{\omega}\frac{S^2\cos\left[\frac{\pi}{2}(1-\varepsilon)\right]}{\cosh\left[\varepsilon\ln(\omega\tau_0)\right] + \sin\left[\frac{\pi}{2}(1-\varepsilon)\right]} + \frac{2(1-S^2)\tau}{1+(\omega\tau)^2}\right] \quad (15)$$

This form of the spectral density function can be further simplified by using the following relationships: $\cosh[\varepsilon\ln(\omega t)] = (1/2)[1 + (\omega\tau)^{2\varepsilon}]/(\omega\tau)^{\varepsilon}$, $\cos[(1-\varepsilon)\pi/2] = \sin(\varepsilon\pi/2)$ and $\sin[(1-\varepsilon)\pi/2] = \cos(\varepsilon\pi/2)$:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\omega^{\varepsilon-1}\tau_0^{\varepsilon}\sin\left(\frac{\pi}{2}\varepsilon\right)}{1+(\omega\tau_0)^{2\varepsilon} + 2(\omega\tau_0)^{\varepsilon}\cos\left(\frac{\pi}{2}\varepsilon\right)} + \frac{(1-S^2)\tau}{1+(\omega\tau)^2}\right] \quad (16)$$

As can be seen from Equation 16 the Cole–Cole spectral density function is not defined at zero frequency, J(0), when ε is less than 1 (due to the $\omega^{\varepsilon-1}$ term). To perform calculations using the Cole–Cole spectral density function, we have assumed that the value of J(0) is equal to the value at J(1). This approximation corresponds to a condition for which there are no interconversion events slower than 1 rad/s. Based on the spectra, this approximation appears to be valid as two or more sets of signals would be observable if

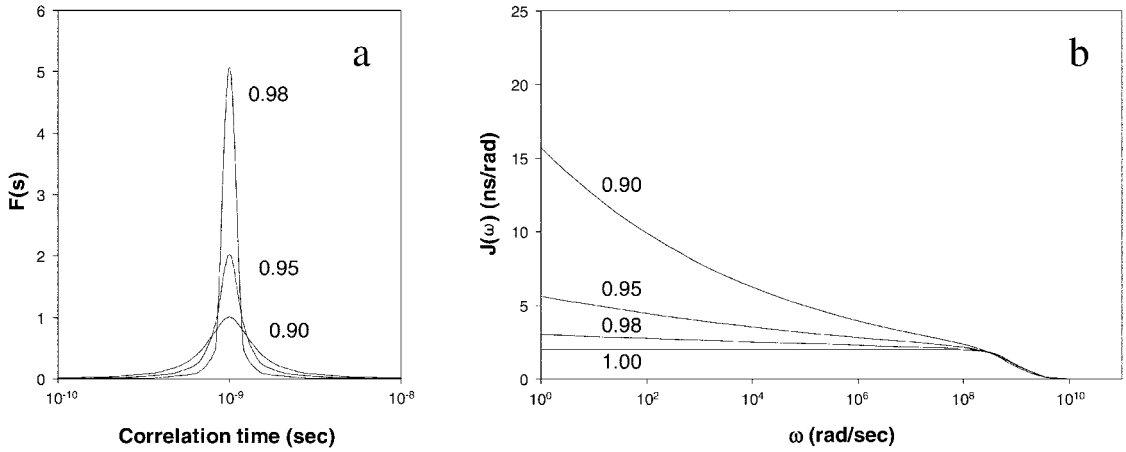

Figure 1. (a) Plots of the Cole–Cole distribution functions shown in Equation 12 with ε values of 0.98, 0.95 and 0.90; and (b) Cole–Cole spectral density functions (Equation 13) with mean correlation times, τ_0, of 1 ns and width of distribution, ε, of 1.00, 0.98, 0.95 and 0.90.

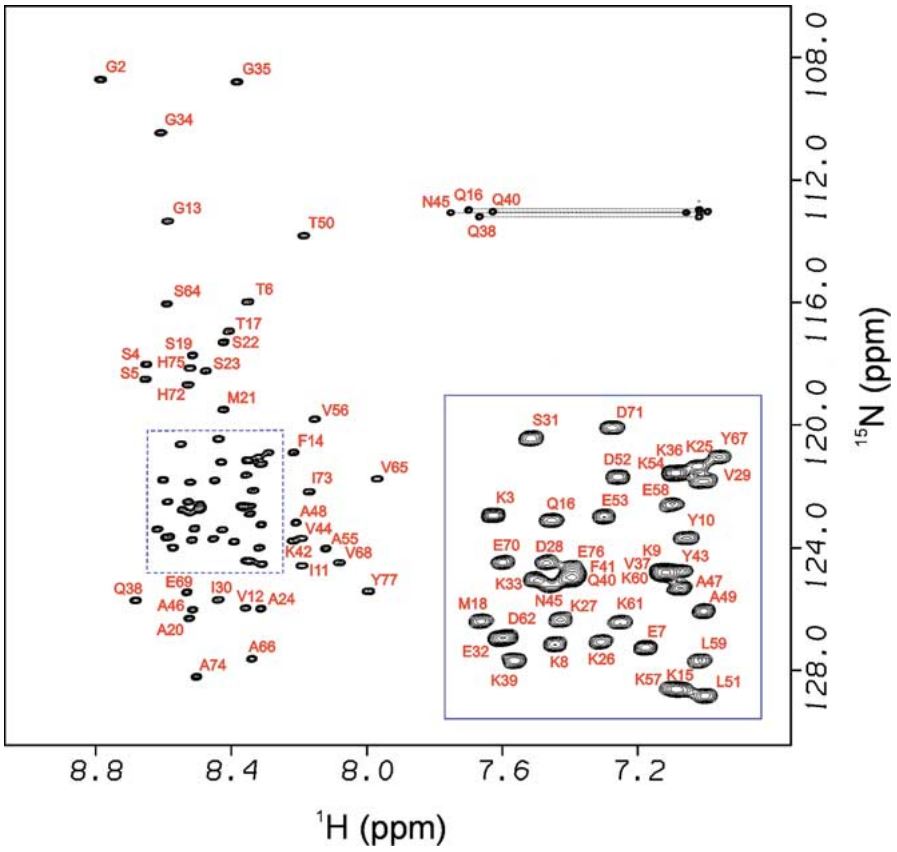

Figure 2. ^{1}H-^{15}N HSQC spectra of uniformly ^{15}N labelled pro-peptide of subtilisin at 9 °C, 0.7 mM, 50 mM sodium phosphate buffer, pH = 6.0. In the inset, the resonance assignments for the central part of the spectrum (dashed rectangle) are presented.

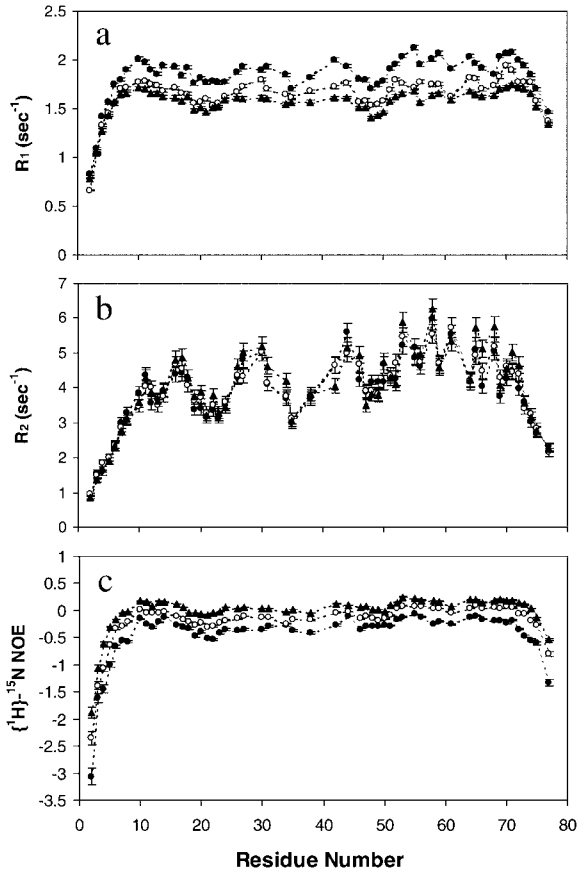

Figure 3. ^{15}N relaxation data R_1 (a), R_2 (b) and $\{^1H\}$-^{15}N NOE (c) of uniformly ^{15}N labelled pro-peptide of subtilisin at 9 °C, 0.7 mM, 50 mM sodium phosphate buffer, pH = 6.0, measured at three magnetic field strengths: 400 MHz (●), 500 MHz (○) and 600 MHz (▲).

interconversion events slower than 1 rad/s existed and the narrow spectral line widths, $\Delta\nu_{1/2}$ = 3–5 Hz, indicate that all processes are much faster than 1 rad/s. To test this approximation and the effect on the model free analysis, the 'zero' frequency was varied from 10000 to 10^{-20}. There were no significant changes in τ_0, S^2 and τ_e; however, the value of ε increased towards 1 when the trial 'zero' frequency moved towards zero. Therefore, the selection of the 'zero' frequency defines the range of ε and acts as an overall scaling factor. Relative differences in ε can be compared within a single protein, and comparisons can be made across different proteins or different solution conditions if the same value of ω is chosen for all the analyses.

Results and discussion

Resonance assignments

Assignments of the ^{1}H and ^{15}N resonances for the propeptide of subtilisin at pH 6.0 and 9 °C (Figure 2) were carried out using a combination of 3D NOESY-HSQC, 3D TOCSY-HSQC and 3D HMQC-NOESY-HSQC experiments. As expected for a natively unfolded protein, the ^{1}H-^{15}N HSQC spectrum displays significant chemical shift degeneracy in both the ^{1}H and ^{15}N dimensions. Based on 265 peaks in the ^{15}N edited TOCSY-HSQC spectrum, 65 spin systems were generated by the Assign module of the Felix-97 program (Molecular Simulation, Inc.). By manual inspection of overlapped peaks we were able to assign an additional 8 spin systems. Even though the ^{15}N edited NOESY-HSQC spectrum contained more than 600 peaks, the low dispersion of ^{1}H chemical shifts and the large number of residues of one type (15 lysines and 10 alanines) allowed only about 25% of the spin systems to be assigned. Therefore the 3D HMQC-NOESY-HSQC experiment, which has substantially broader signal dispersion, needed to be included in the analysis. Based on the NH–NH NOE cross peaks of (i, i + 1) residues from the HMQC-NOESY-HSQC experiment together with the Hα–NH and H(side chain)–NH NOE cross peaks of (i, i + 1) residues from the NOESY-HSQC experiment, the ^{1}H and ^{15}N assignments have been readily completed for 75 residues of the propeptide (Figure 2). The only residues that could not be assigned were Pro63 and the N-terminal Ala1. For Ser64, two sets of signals in a 9:1 ratio were observed. The close proximity of Pro63 suggests that the two signals for Ser64 correspond to the *trans*- and *cis*-isomers of the Asp62-Pro63 bond, with the more populated form corresponding to the *trans*-prolyl conformation (Grathwohl and Wüthrich, 1976).

^{15}*N relaxation data*

^{15}N relaxation experiments have been used to probe the dynamical behavior of the unfolded pro-peptide. ^{15}N R_1, R_2 and heteronuclear $\{^1H\}$-^{15}N NOE values have been measured at 9 °C, pH 6.0 at three magnetic field strengths (400, 500 and 600 MHz) for 55 residues (Figure 3). Severe resonance overlap did not allow accurate measurement of relaxation rates for the remaining residues. The R_1, R_2 and $\{^1H\}$-^{15}N NOE data show the expected magnetic field dependence, with R_1 and $\{^1H\}$-^{15}N NOE values increasing, and R_2 values decreasing with increasing magnetic field strength. R_2 rate constants were measured at two con-

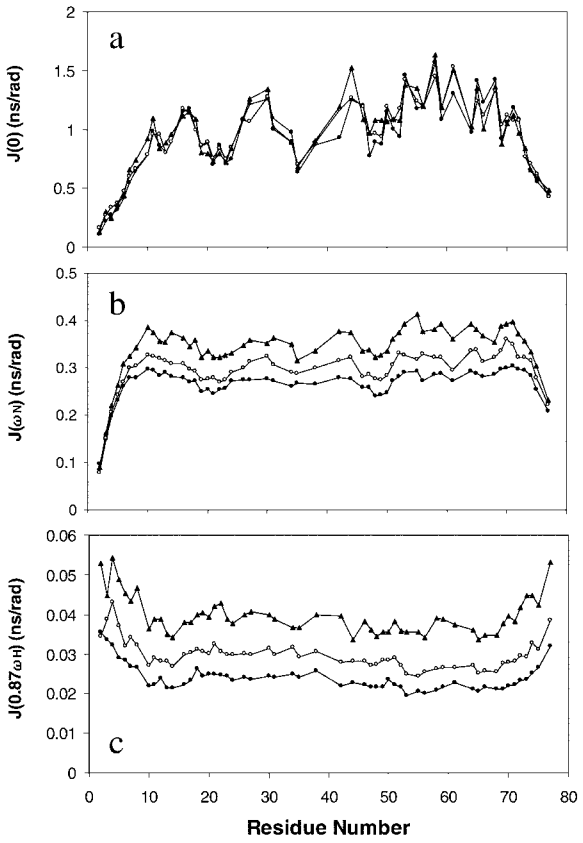

Figure 4. Reduced spectral density mapping analysis. Spectral density amplitudes J(0) (a), J(ω_N) (b) and J($0.87\omega_H$) (c) for the pro-peptide of subtilisin at 9 °C derived from the relaxation data R_1, R_2 and {^{1}H}-^{15}N NOE acquired at three magnetic field strengths shown in Figure 3.

centrations including 0.7 and 0.35 mM. The results show that the R_2 relaxation rates are unchanged as a function of concentration, suggesting that the protein is not aggregating under the experimental conditions. The average R_1 and R_2 values are 1.7 s^{-1} and 4.7 s^{-1} at 500 MHz which are similar to those observed in random coil and in measurements of other unfolded or denatured proteins (Farrow et al., 1995; Frank et al., 1995; Schwalbe et al., 1997; Penkett et al., 1998). Overall, the R_1 and {^{1}H}-^{15}N NOE curves are rather uniform along the sequence, whereas the pattern of R_2 rates is much more variable and heterogeneous, ranging from 3 to 6 s^{-1}. All relaxation data have a bell shape dependence along the sequence, indicating that both termini have different flexibility compared with the internal residues. Moreover, the residues at the two termini have slightly different relaxation properties. At the C-terminus the average relaxation rates are reached by the 4–5th residue, whereas at the N-terminus the average rates are reached only by the 8–10th residue. Qualitatively this indicates more flexibility at the N-terminus relative to the C-terminus.

Reduced spectral density mapping

The reduced spectral density mapping approach has been previously used to analyze relaxation data for both folded and unfolded proteins (Farrow et al., 1995; Buck et al., 1996; Meekhof and Freund, 1999; Eliezer et al., 2000). It allows sampling of the spectral density at zero, ω_N and $0.87\omega_H$ frequencies without making an assumption about a model of relaxation. Reduced spectral density values for unfolded pro-peptide at pH 6.0 and 9 °C are shown in Figure 4 at three different magnetic field strengths. The field dependence of the spectral density values follows the order $J(0) \gg J^{400}(\omega_N) > J^{500}(\omega_N) > J^{600}(\omega_N) \gg J^{400}(0.87\omega_H) > J^{500}(0.87\omega_H) > J^{600}(0.87\omega_H)$, and shows the expected field dependence for J(ω_N) and J($0.87\omega_H$). The field dependence for J(0) is complicated and may arise from chemical exchange contributions to R_2. Exchange contributions to R_2, or slow micro- to millisecond motions, are reflected in J(0) spectral density values as an increase in J(0) values, with J(0) expected to go as the square of the magnetic field strength assuming two-site exchange and equal populations (Equation 6). The criteria that were used to establish whether specific residues were field dependent were the following. If J(0) values for specific residues increased as expected from 400 to 600 MHz, then these were considered to be undergoing motions on the millisecond timescale. The spectral density values for J(0) are generally small, on the order of 0.5×10^{-9} to 1.5×10^{-9} rad/s for the central region of the protein. Overall, it can be seen that there is minimal field dependence in J(0), suggesting that most of the residues are not undergoing conformational exchange on the slow millisecond timescale. Comparison of the J(0) amplitudes for each residue as a function of two field strengths (400 and 600 MHz) shows that 50 out of 55 residues do not reveal any field dependence. For the remaining residues Ser22, Ala46, Glu53 and His72 the field dependent increment ($J^{600}(0) - J^{400}(0)$) was less than 10% of the value of $J^{600}(0)$, and for Ala66 it was about 20% of the $J^{600}(0)$ value, suggesting that even for these residues the R_{ex} is small. These data indicate that the variability and increased values of J(0) do not appear to derive from slow timescale motions, as evidenced from the lack of field dependence in J(0).

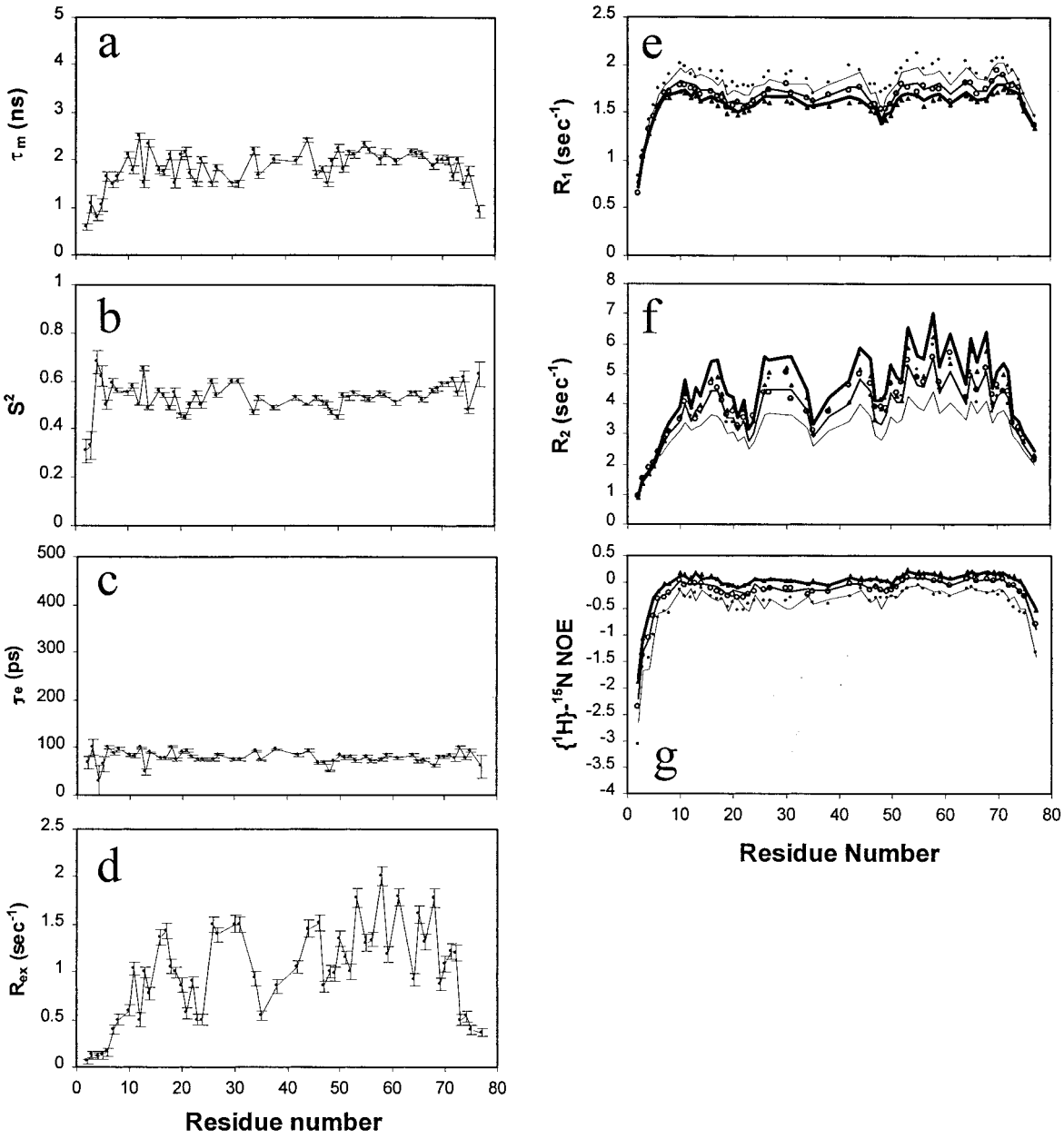

Figure 5. Model free analysis (*Model 5*) of the multiple field ^{15}N relaxation data of pro-peptide. Residue specific correlation times, τ_m (a), order parameter S^2 (b), internal correlation times τ_e (c) and R_{ex} (d) were calculated using the Model Free 3.1 program from the relaxation data of pro-peptide measured at three magnetic field strengths shown in Figure 3. Experimental values of R_1 (e), R_2 (f) and $\{^1H\}$-^{15}N NOE (g) are shown by (●) 400 MHz; (○) 500 MHz and (▲) 600 MHz. Model free fittings of experimental R_1, R_2 and $\{^1H\}$-^{15}N NOE are shown as a thin line (400 MHz), a medium line (500 MHz) and a thick line (600 MHz). Fitting of R_1 and $\{^1H\}$-^{15}N NOE data was satisfactory, whereas the R_2 field dependence was largely overestimated, as described in the text.

Model free analysis

The model free (MF) analysis of Lipari and Szabo (1982a, b) is frequently used to interpret relaxation data for folded proteins and has also been employed for the analysis of unfolded proteins (Alexandrescu and Shortle, 1994; van Mierlo et al., 1994; Farrow et al., 1995, 1997; Frank et al., 1995; Buck et al., 1996). A number of models can be used to analyze the relaxation data of folded and unfolded proteins, including those with a single overall correlation time, those that consider fast and slow internal motions, those that use a local correlation time for each residue and those that include a conformational exchange term R_{ex}. Relaxation data of pro-peptide were analyzed by five different models, characterized by different sets of variable parameters in Equations 10 and 11:

Model 1: τ_m (global), S^2, τ_e
Model 2: τ_m (global), S^2, τ_e, R_{ex}
Model 3: τ_m (global), S^2, S_f^2, τ_e, R_{ex}
Model 4: τ_m (local), S^2, τ_e
Model 5: τ_m (local), S^2, τ_e, R_{ex}

Models 1–3 (Supplementary material) estimated global τ_m values of 2.25, 1.99 and 2.20 ns, which were close to the average values of τ_m found by *Models 4* and *5* (2.13±0.46 and 1.81±0.41 ns, respectively). The τ_e correlation times in all models, except *Model 3*, were in the range of 50–100 ps. The reason that *Model 3* gave different results was due to the intrinsic correlation of the S_s^2, S_f^2 and τ_e parameters, which subsequently resulted in their much larger estimated errors (Supplementary material). It is noteworthy that the *Model 3* analysis showed that even with three magnetic field strengths, the extension of the model to three types of motions (or the use of three Lorentzians in the spectral density language) results in less accurate estimation of the correlation times and order parameters. The generalized order parameter S^2, found in all calculations, ranged from 0.4 to 0.6; however, in the models with fewer varied parameters (those with τ_m global) the sequence dependence of S^2 turned out to be more smooth, closely resembling the shapes of R_1 and $\{^1H\}$-^{15}N NOE data (Supplementary material).

The model that provided the lowest overall R-factor (Table 1) was *Model 5* (Figure 5). It can be seen that the fit to the $\{^1H\}$-^{15}N NOE (Figure 5g) and R_1 (Figure 5e) values was good but the fit to R_2 (Figure 5f) is not adequate and the R_2 field dependence is overestimated. Specifically, the R_2 component of the R-factor remains rather high, 0.14, whereas the R_1 R-factor is only 0.026. These results of the MF analysis are consistent with the observation of increased J(0) values shown in the section above. To fit the data using *Model 5*, R_{ex} terms are incorporated into the model free analysis (Figure 5d) in order to improve the fit to the R_2 data, thereby suggesting that the protein is undergoing pervasive motion on the millisecond timescale. However, we have seen through the lack of magnetic field dependence of J(0) values and through the overestimation of the field dependence in model free (*Model 5*) that R_{ex} terms applied to specific residues may not be meaningful or appropriate compensation for the higher R_2 values.

In the models where the R_{ex} term was not included, R_2 values were significantly underestimated, whereas in the models that included the R_{ex} terms, the R_2 field dependence was overestimated. To understand these results one has also to consider the properties of the minimization algorithm of the Model Free program. The sensitivity of the minimization procedure of the Model Free program is proportional to the accuracy of experimental data (A. Palmer, Model Free, V 3.1). In case of the pro-peptide the accuracy of R_2 values was approximately five times lower than that of R_1. Therefore the program was more sensitive to the R_1 and $\{^1H\}$-^{15}N NOE field dependence and less sensitive to R_2, resulting in a large discrepancy in R_2. One can force the minimizer to reduce the R_{ex} term by deliberately increasing the R_2 accuracy; however, that in turn creates unrealistically large discrepancies between experimental and theoretical $\{^1H\}$-^{15}N NOE values (Supplementary material).

Cole–Cole model free analysis

A new approach to the analysis of relaxation data of unfolded proteins was recently suggested (Buevich and Baum, 1999), and incorporates a Cole–Cole distribution of nanosecond correlation times into the model free approach (Equation 16). Two types of models, analogous to *Models 4 and 5* of the standard model free approach, were considered:

CC-Model 1: τ_0 (local), ε, S^2, τ_e
CC-Model 2: τ_0 (local), ε, S^2, τ_e, R_{ex}

CC-Model 2 provided the best R-factor (Table 1, Figure 6) relative to *CC-Model 1*, and provided a much better fit to the experimental data than *Model 5* of the conventional model free analysis (Figures 6f, 6g, 6h). Thus the R-factors of the CC-MF analysis were approximately three times lower than the best of the conventional model free analysis, due mainly to the

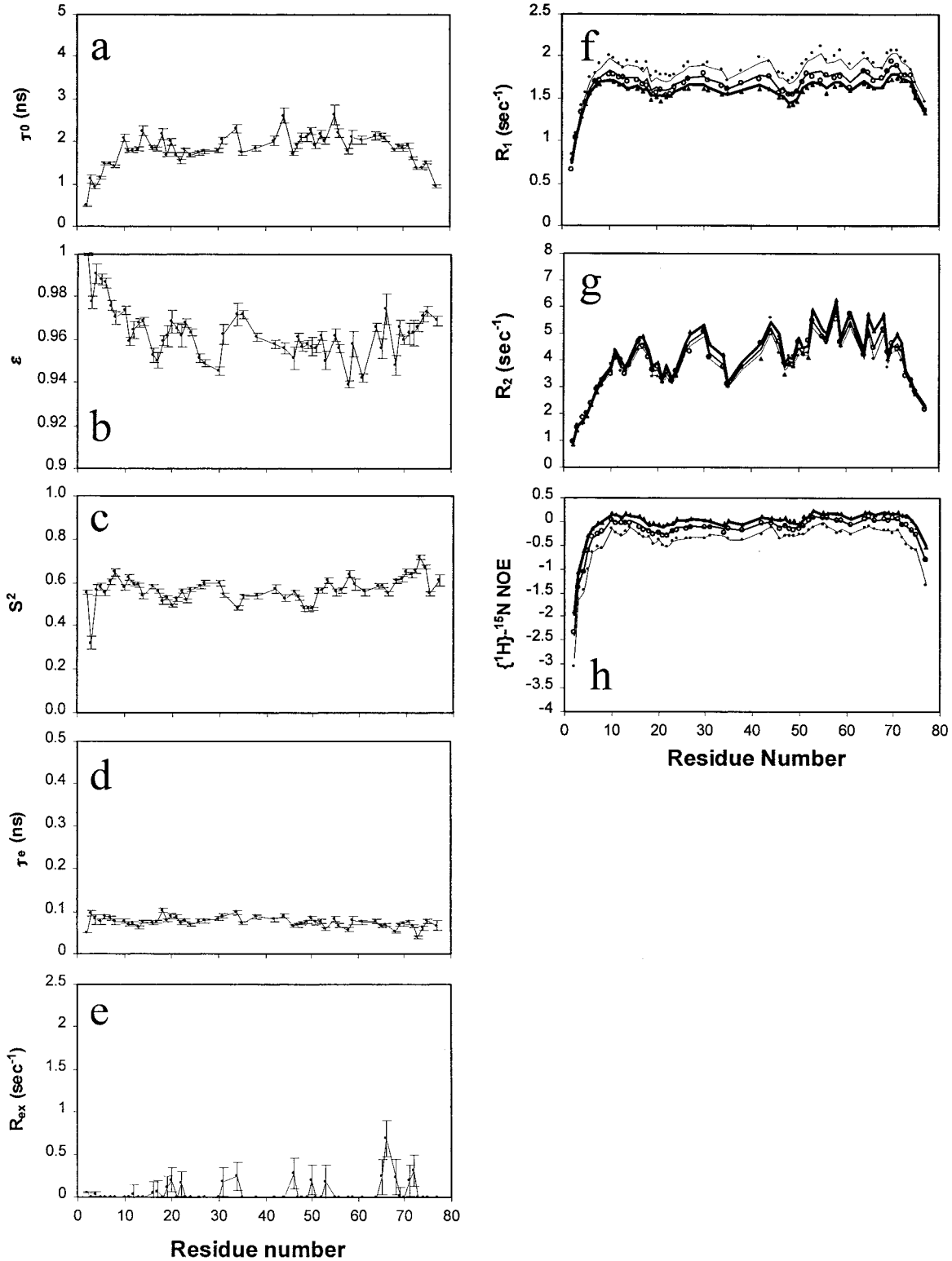

Figure 6. Cole–Cole model free analysis (*CC-Model 2*) of the multiple field relaxation data of pro-peptide. Residue-specific mean correlation times, τ_0, (a), width of the distribution ε (b), order parameter S^2 (c), internal correlation times τ_e (d) and R_{ex} (e) were obtained by least-squares fit to the nine sets of R_1, R_2 and $\{^1H\}$-^{15}N NOE relaxation data of pro-peptide measured at three field strengths shown in Figure 3. Experimental values of R_1 (f), R_2 (g) and $\{^1H\}$-^{15}N NOE (h) are shown by (●) 400 MHz; (○) 500 MHz and (▲) 600 MHz. Cole–Cole model free fittings of experimental R_1, R_2 and $\{^1H\}$-^{15}N NOE are shown as thin (400 MHz), medium (500 MHz) and thick lines (600 MHz).

Table 1. R-factor values for the model free and Cole–Cole model free analysis of ^{15}N R_1, R_2, and {^{1}H}-^{15}N NOE relaxation data of uniformly ^{15}N labelled pro-peptide of subtilisin at pH 6.0 measured at three magnetic field strengths (400, 500 and 600 MHz)[a]

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*CC-model 1*	*CC-model 2*
R-factor (R_1)	0.025	0.030	0.026	0.026	0.032	0.024	0.023
R-factor (R_2)	0.434	0.134	0.131	0.435	0.139	0.044	0.041
R-factor (NOE)	0.283	0.226	0.113	0.226	0.141	0.070	0.061
R-factor (R_1+R_2+NOE)	0.399	0.126	0.122	0.400	0.129	0.042	0.039

[a]*Model 1*: model free analysis with τ_m (global), S^2, τ_e varied parameters; *Model 2*: τ_m (global), S^2, τ_e, R_{ex}; *Model 3*: τ_m (global), S^2, S_f^2, τ_e, R_{ex}; *Model 4*: τ_m (residue specific), S^2, τ_e; *Model 5*: τ_m (residue specific), S^2, τ_e, R_{ex}; *CC-model 1*: model free analysis with Cole–Cole distribution function and τ_0 (residue specific), ε, S^2, τ_e varied parameters; *CC-model 2*: model free analysis with Cole–Cole distribution function and τ_0 (residue specific), ε, S^2, τ_e and R_{ex} varied parameters.

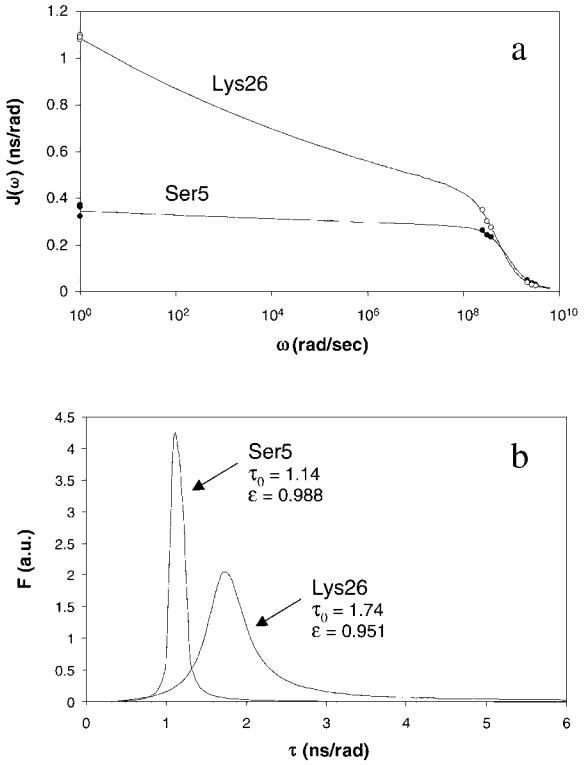

Figure 7. Comparison of Cole–Cole model free analysis and spectral density mapping analysis. Data for two representative residues, Ser5 and Lys26, are shown to illustrate how the spectral density values derived by spectral density mapping (Ser5 (●) and Lys26 (○)), in particular J(0) amplitudes, can be fit using the Cole–Cole model free approach (a). Distributions of correlation times applied in Cole–Cole model free analysis for Ser5 and Lys26 are shown in (b).

substantial reduction of the discrepancies in R_2 estimations (Table 1). The average values of τ_0, S^2 and τ_e were 1.81 ns, 0.57 and 74 ps, respectively (Figure 6a, 6b, 6c). These values turned out to be very similar to those found by *Model 5*, suggesting that t_0, S^2 and τ_e are less sensitive to the distribution function than R_{ex} (Figure 6e). In contrast to *Model 5*, the CC-model free analysis shows that only 19 of 55 residues have small R_{ex} terms with values no larger than 0.7 s^{-1}. In the spectral density mapping approach only 5 residues showed significant field dependence. One may expect that the number of residues showing significant field dependence in the spectral density analysis and the number of residues requiring an R_{ex} term in the Cole–Cole analysis would be similar. However, considering the small values of the R_{ex} term and the magnitude of error associated with them, and the weak criteria for field dependence applied to the spectral density analysis, these two approaches give similar conclusions about the range of millisecond timescale motions in PPS. In summary, the large R_2 rates that could not be fit adequately using the model free approach may be accounted for by a distribution of correlation times on the nanosecond timescale. This suggests that the major component of the R_2 variability in unfolded proteins arises from a distribution of correlation times on the nanosecond timescale described by the parameter ε.

Relation of Cole–Cole model free analysis and spectral density mapping

Simulations of the Cole–Cole spectral density functions (Figure 1b) show that distributions of correlation times on the nanosecond timescale lead to increased values of J(0) relative to Lorentzian spectral density functions. The shape of the Cole–Cole spectral den-

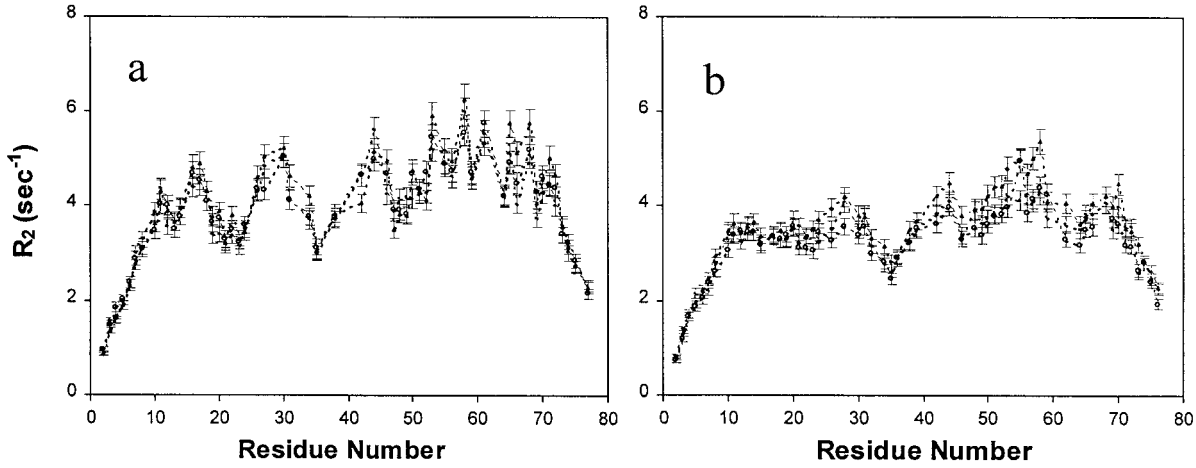

Figure 8. R_2 relaxation data for pro-peptide of subtilisin measured at 9 °C at three magnetic field strengths at pH 6.0 (a) and pH 3.0 (b) (400 MHz (●), 500 MHz (○) and 600 MHz (▲)). R_1 and $\{^1H\}$-^{15}N NOE at pH 3.0 were similar to those at pH 6.0 and are available as Supplementary material.

sity function is sensitive to the width of distribution of correlation times and its initial slope is directly proportional to the width of the distribution. Representative computer simulations are performed for two residues, Ser5 and Lys26, using the Cole–Cole spectral density function to illustrate how the spectral density values derived by spectral density mapping, in particular J(0), can be fit using this function (Figure 7a). Two different distributions of correlation times, a very narrow distribution, $\varepsilon = 0.99$ for Ser5, and a relatively broad one, $\varepsilon = 0.95$ for Lys26, are shown (Figure 7b). Both residues have slightly different local correlation times as determined from the CC-MF analysis. The example of Ser5 with a narrow distribution function shows that at low frequencies the spectral density is nearly constant, which closely resembles the situation typically observed for folded proteins where the spectral density function has a Lorentzian shape. In contrast, the spectral density values for Lys26 indicate that a distribution of correlation times leads to an increase in J(0) and a non-Lorentzian shape. Therefore the Lys26 data can be fit with a Cole–Cole spectral density function only by optimization of the width of the distribution of correlation times. An attempt to fit the data for Lys26 with the Lorentzian spectral density function assumed in the model free analysis will result in large discrepancies in J(0) or in $J(0.87\omega_H)$ amplitudes. Incorporation of a distribution of correlation times on the nanosecond timescale improves the fit to the increased values of J(0) obtained through the spectral density mapping and allows the model free approach to be used with significantly improved fits to R_2.

pH Dependence of relaxation data

To assess the effect of electrostatic interactions in the dynamics of pro-peptide, ^{15}N relaxation data for pro-peptide at pH 3.0 were measured at three magnetic fields (400, 500 and 600 MHz) at 9 °C. Figure 8 shows the comparison of the R_2 values obtained at pH 6.0 (Figure 8a) and pH 3.0 (Figure 8b) (R_1 and $\{^1H\}$-^{15}N NOE relaxation data are presented in the supplementary material). It can be seen that lowering the pH had a large effect on R_2 rates, whereas R_1 and $\{^1H\}$-^{15}N NOE data were nearly identical (Supplementary material). In particular, the R_2 values are generally smoother at low pH than at high pH and do not exhibit the variability seen at high pH. Comparison of relaxation data of pro-peptide at pH 3.0 and 6.0 show that the largest effect of the pH change arises in the parameter ε, whereas S^2, τ_0, τ_e and R_{ex} remain similar for the majority of the residues (Figure 9). It is striking to observe that the heterogeneity of ε diminishes dramatically upon protonation of the acidic residues. For all residues at pH 3.0, ε values are increased, indicating a decreased width of distributions on the nanosecond time scale. In particular, large changes were observed for the residues with polar or charged side chains, such as Glu7-Lys8; Lys15-Gln16-Thr17; Lys25-Lys26-Lys27; Glu58. Moreover, the maximum increase in ε was found to be in good agreement with the position of charged residues, such as the cluster of residues at positions 8–18, 25–32 and 50–62. For example, the cluster at position 25 contains the residues KKKDVISEK, a highly charged region for which the value of ε increased significantly. However, the effect

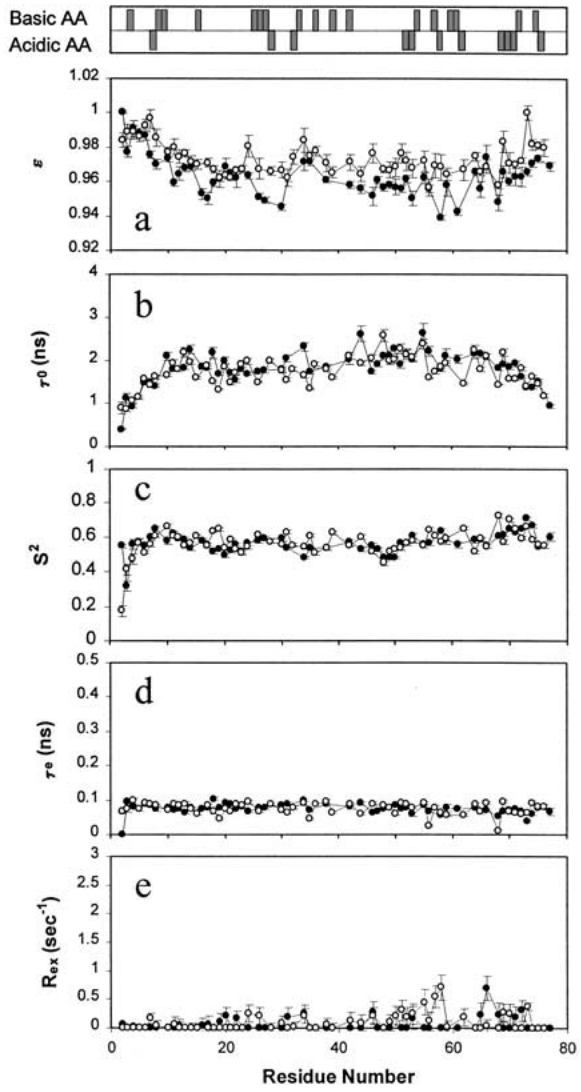

Figure 9. Comparison of the Cole–Cole model free analyses (*CC-Model 2*) of relaxation data of pro-peptide at pH 6.0 (●) and 3.0 (○). Residue specific width of the distributions ε (a), mean correlation times τ_0, (b), order parameters S^2, (c), internal correlation times τ_e, (d), and R_{ex} (e) were obtained by least-squares fitting to the two sets of relaxation data of pro-peptide at pH 6.0 and 3.0. Each of these sets contains nine subsets of R_1, R_2 and $\{^1H\}$-^{15}N NOE relaxation data measured at three field strengths (Figure 3 and supplementary material). Positions of basic and acidic residues in pro-peptide are shown at the top of the figure by positive and negative bars, respectively.

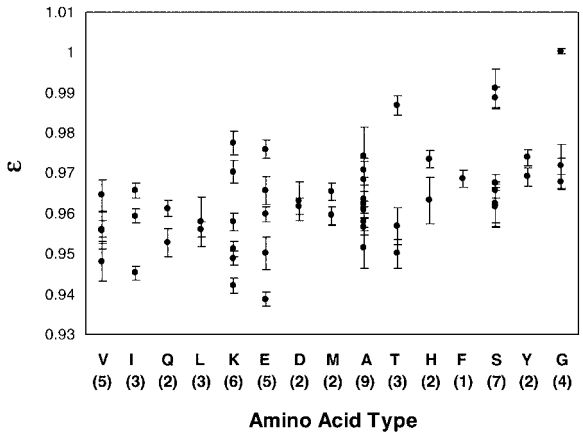

Figure 10. Specific values of ε and respective error bars for pro-peptide at pH 6.0 as a function of amino acid type. The number of times the amino acid appears in the pro-peptide sequence is shown in parentheses.

of pH can also be found among some hydrophobic residues, such as Ile11, Ala24, Ile30. This suggests that electrostatic and hydrophobic interactions are not completely independent, but rather their co-operative effect defines the complex dynamic behaviour of the unfolded protein chain.

Biological implications of the Cole–Cole model free analysis

The width of the distribution of correlation times, ε, as a function of amino acid sequence is likely to arise from the diversity of motions that every residue participates in and reflects the heterogeneity of interactions with other residues in the sequence. A plot of ε versus residue type (Figure 10) shows that each amino acid can have a range of ε values. The range of values of ε, within a specific amino acid type, suggests that ε is not only defined by amino acid type but also by local sequence and interactions with neighboring residues. As has been described for polymers (Bovey and Mirau, 1996), their motions can be classified into rapid short range processes and slow longer range processes, suggesting that the greater the range in correlation times, the wider the distribution (Heatley and Begum, 1976). The source of the distribution for unfolded proteins might arise from a diversity of interresidue interactions including electrostatic, dipolar and hydrophobic. Therefore a residue without side chains or with less prominent properties is expected to have narrower width of distribution. Accordingly, the narrowest distributions are observed for small uncharged residues, like Gly34 and Gly35, and also for the termi-

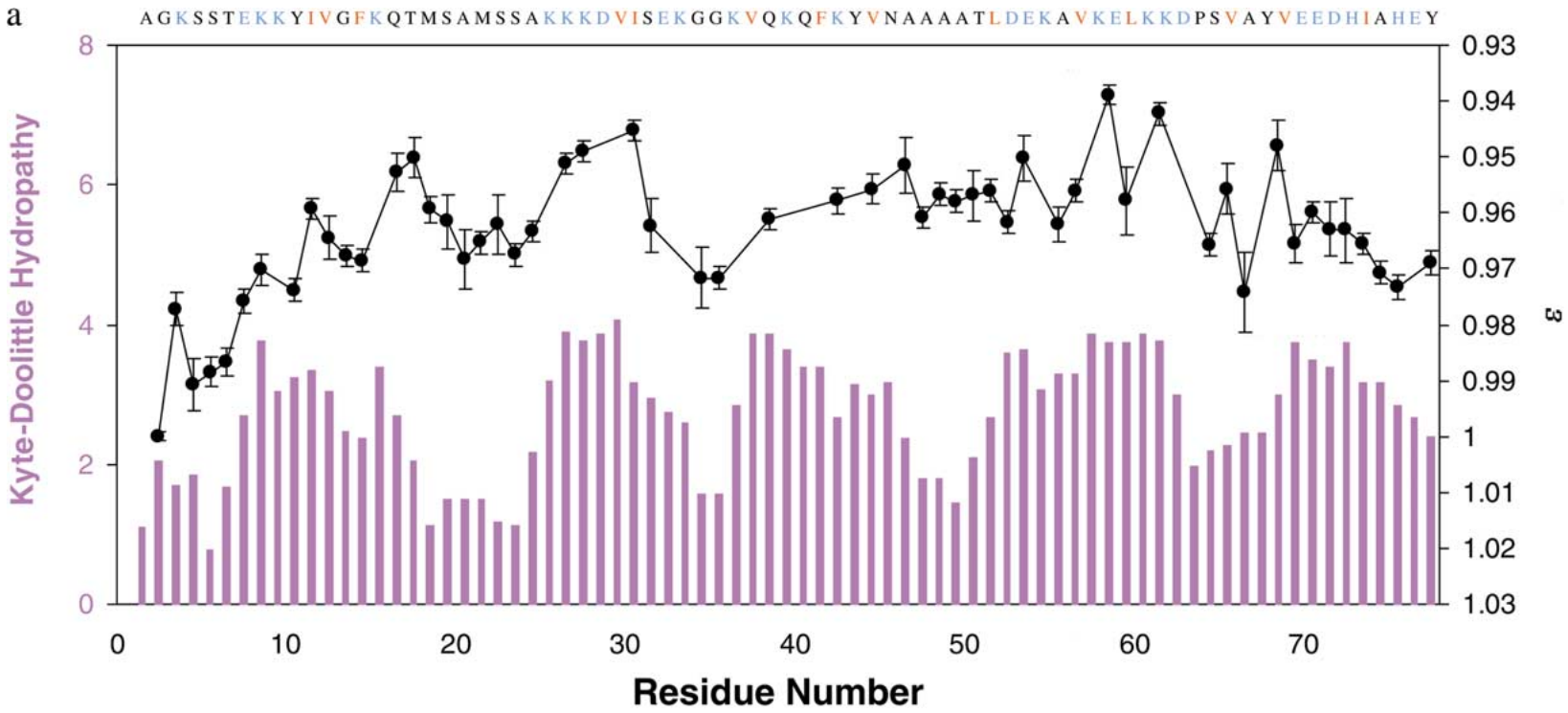

Figure 11. Correlation of Kyte–Doolittle hydropathy (KDH) and the width of the distribution of correlation times ε for pro-peptide of subtilisin at pH = 6.0. (a) KDH is presented by a bar graph and ε is a line graph; the values of ε are presented on a reverse vertical scale to simplify the comparison with the hydropathic property. (b) Correlation plot showing ε versus KDH. The KDH of pro-peptide was smoothed on the basis of a three-residue window. The power spectrum of KDH was used rather than absolute values as the power spectrum more adequately represents an integral hydrophobic/hydrophilic property of amino acid side chains. This prevents an annihilation of the hydropathic effects upon smoothing when adjacent residues have hydropathy of different signs. The correlation between ε and KDH, seen in both plots, suggests that ε may be related to the inter-residual electrostatic and hydrophobic interactions. The high proportion of charged residues within the pro-peptide sequence (charged residues are coded in blue and hydrophobic ones in red) also suggests that ε may be defined mostly by the diversity of electrostatic inter-residual interactions.

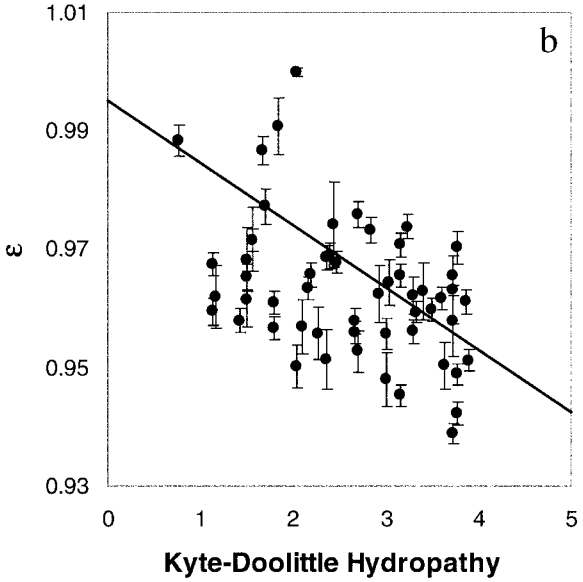

Figure 11. (continued).

nal residues which have statistically less contact than intra-chain residues.

The relationship between the width of the distribution function, ε, and the amino acid composition of pro-peptide can be evaluated using the Kyte–Doolittle hydropathy (KDH) scale (Kyte and Doolittle, 1982). Figure 11a shows a plot of ε superimposed on KDH values as a function of sequence. In general, it can be seen that regions with smaller values of ε, or with larger nanosecond timescale distributions, correspond to the charged/hydrophobic regions of the pro-peptide. The local maxima in ε correlate with regions that do not contain charged/hydrophobic residues, and the asymmetry of the dynamics in the two terminal ends is also reflected in their Kyte–Doolittle values. To establish the correlation between KDH and ε more quantitatively, Figure 11b shows a plot of ε versus KDH directly. The anticorrelation across the entire sequence is not very strong (-0.43); however, certain highly charged/hydrophobic regions, in particular residues 19–35, have a higher anticorrelation coefficient of -0.85. These data are consistent with those observed in the literature where a number of unfolded proteins have J(0) or R_2 variability that was found to be correlated with clustering of charged and hydrophobic residues (Penkett et al., 1998; Meekhof and Freund, 1999).

The relationship of the Kyte–Doolittle hydropathy with ε (Figure 11), along with the pH dependent changes in ε (Figure 9), suggest that the distribution of nanosecond time scale dynamics in the unfolded PPS derives primarily from electrostatic interactions. Examination of the amino acid sequence of PPS shows that only approximately 18% of the protein contains hydrophobic residues and 36% of the protein contains charged residues. Therefore the major contribution to the Kyte–Doolittle hydropathy derives from the charged residues in the sequence. The correlation of ε with Kyte–Doolittle hydropathy and the pH dependent changes in ε indicate that charge interactions between residues may be important for inducing correlated motions in unfolded PPS and that the nanosecond time scale distributions are very sensitive to the ionization state of the protein. Specifically, the unfolded protein samples a larger distribution of states at neutral pH than at low pH arising from, for example, the formation of hydrogen bonds or ion pairs. In the language of polymer dynamics, the neutral pH form of propeptide is undergoing more correlated motions within the sequence than the low pH form. At low pH, when acidic residues are protonated, the protein exhibits fewer interresidue interactions along the sequence.

Describing site specific residues of unfolded proteins in terms of distributions of correlation times on the nanosecond timescale will help shed new light on the nature of the unfolded state by providing a new framework for thinking about the timescales of motions and the types of correlated motions that are prominent in unfolded proteins. The pH dependent results suggest the existence of variations in the dynamics of unfolded proteins as a function of pH. In particular, the results demonstrate that nanosecond time scale motions are important in defining the unfolded state of PPS and that electrostatic interactions may create the correlated motions on these timescales.

Acknowledgements

We thank Dr. Art Palmer for stimulating discussions and for the use of the Model Free program, and Dr. Lewis Kay for providing pulse sequences of 3D experiments. This work was supported by NIH grant GM-45302 (J.B.).

References

Abragam, A. (1961) *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford.

Alexandrescu, A.T. and Shortle, D. (1994) *J. Mol. Biol.*, **242**, 527–546.

Arcus, V.L., Vuilleumier, S., Freund, S.M., Bycroft, M. and Fersht, A.R. (1995) *J. Mol. Biol.*, **254**, 305–321.
Bovey, F.A. and Mirau, P.A. (1996) *NMR of Polymers*, Academic Press, San Diego, CA.
Brutscher, B., Bruschweiler, R. and Ernst, R.R. (1997) *Biochemistry*, **34**, 5904–5912.
Buck, M., Schwalbe, H. and Dobson, C.M. (1996) *J. Mol. Biol.*, **257**, 669–683.
Buevich, A.V. and Baum, J. (1999) *J. Am. Chem. Soc.*, **121**, 8671–8672.
Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C. and Gronenborn, A.M. (1990) *J. Am. Chem. Soc.*, **112**, 4989–4991.
Cohen, F.E. and Prusiner, S.B. (1998) *Annu. Rev. Biochem.*, **67**, 793–819.
Cole, K.S. and Cole, R.H. (1941) *J. Chem. Phys.*, **9**, 341–351.
Dill, K.A. and Shortle, D. (1991) *Annu. Rev. Biochem.*, **60**, 795–825.
Dobson, C.M. and Hore, P.J. (1998) *Nat. Struct. Biol.*, **5** (Suppl.), 504–507.
Dobson, C.M., Sali, A. and Karplus, M. (1998) *Angew. Chem. Int. Ed. Engl.*, **37**, 868–893.
Dyson, H.J. and Wright, P.E. (1998) *Nat. Struct. Biol.*, **5** (Suppl.), 499–503.
Eder, J., Rheinnecker, M. and Fersht, A.R. (1993) *J. Mol. Biol.*, **233**, 293–304.
Eliezer, D., Chung, J., Dyson, H.J. and Wright, P.E. (2000) *Biochemistry*, **39**, 2894–2901.
Eliezer, D., Yao, J., Dyson, H.J. and Wright, P.E. (1998) *Nat. Struct. Biol.*, **5**, 148–155.
Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) *Biochemistry*, **33**, 5984–6003.
Farrow, N.A., Zhang, O., Forman-Kay, J.D. and Kay, L.E. (1995a) *Biochemistry*, **34**, 868–878.
Farrow, N.A., Zhang, O., Forman-Kay, J.D. and Kay, L.E. (1997) *Biochemistry*, **36**, 2390–2402.
Farrow, N.A., Zhang, O., Szabo, A., Torchia, D.A. and Kay, L.E. (1995b) *J. Biomol. NMR*, **6**, 153–162.
Frank, M.K., Clore, G.M. and Gronenborn, A.M. (1995) *Protein Sci.*, **4**, 2605–2615.
Frenkiel, T., Bauer, C., Carr, M.D., Birdsall, B. and Feeney, J. (1990) *J. Magn. Reson.*, **90**, 420–425.
Fuoss, R.M. and Kirkwood, J.G. (1941) *J. Am. Chem. Soc.*, **63**, 385–394.
Grathwohl, C. and Wüthrich, K. (1976) *Biopolymers*, **15**, 2025–2041.
Heatley, F. and Begum, A. (1976) *Polymer*, **17**, 399–408.
Hiyama, Y., Niu, C.-H., Silverton, J.V., Bavoso, A. and Torchia, D.A. (1988) *J. Am. Chem. Soc.*, **110**, 2378–2383.
Ikemura, H., Takagi, H. and Inouye, M. (1987) *J. Biol. Chem.*, **262**, 7859–7864.
Ikura, M., Bax, A., Clore, G.M. and Gronenborn, A.M. (1990) *J. Am. Chem. Soc.*, **112**, 9020–9022.
Ishima, R. and Nagayama, K. (1995) *J. Magn. Reson.*, **B108**, 73–76.
Ishima, R. and Torchia, D. (1999) *J. Biomol. NMR*, **14**, 369–372.
Kay, L.E. (1998) *Nat. Struct. Biol.*, **5** (Suppl.), 513–517.
Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10665.
Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.
Li, Y. and Inouye, M. (1994) *J. Biol. Chem.*, **269**, 4169–4174.
Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc.*, **104**, 4546–4559.
Lipari, G. and Szabo, A. (1982b) *J. Am. Chem. Soc.*, **104**, 4559–4570.
Meekhof, A.E. and Freund, S.M. (1999) *J. Mol. Biol.*, **286**, 579–592.
Neri, D., Billeter, M., Wider, G. and Wüthrich, K. (1992) *Science*, **257**, 1559–1563.
Palmer, A.G. (1993) *Curr. Opin. Biotechnol.*, **4**, 385–391.
Palmer, A.G., Williams, J. and McDermott, A. (1996) *J. Phys. Chem.*, **100**, 13293–13310.
Peng, J.W. and Wagner, G. (1992) *J. Magn. Reson.*, **98**, 308–332.
Peng, J.W. and Wagner, G. (1995) *Biochemistry*, **34**, 16733–16752.
Penkett, C.J., Redfield, C., Jones, J.A., Dodd, I., Hubbard, J., Smith, R.A., Smith, L.J. and Dobson, C.M. (1998) *Biochemistry*, **37**, 17054–17067.
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) *Biochemistry*, **36**, 8977–8991.
Shaefer, J. (1973) *Macromolecules*, **6**, 882–888.
Shinde, U. and Inouye, M. (1995) *J. Mol. Biol.*, **252**, 25–30.
Shinde, U., Li, Y., Chatterjee, S. and Inouye, M. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 6924–6928.
Shortle, D. (1996) *Curr. Opin. Struct. Biol.*, **6**, 24–30.
Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) *Proteins Struct. Funct. Genet.*, **41**, 415–427.
van Mierlo, C.P., Kemmink, J., Neuhaus, D., Darby, N.J. and Creighton, T.E. (1994) *J. Mol. Biol.*, **235**, 1044–1061.
Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, **6**, 135–140.
Wong, K.B., Freund, S.M. and Fersht, A.R. (1996) *J. Mol. Biol.*, **259**, 805–818.
Wright, P.E. and Dyson, H.J. (1999) *J. Mol. Biol.*, **293**, 321–331.
Zhang, O. and Forman-Kay, J.D. (1995) *Biochemistry*, **34**, 6784–6794.
Zhu, X.L., Ohta, Y., Jordan, F. and Inouye, M. (1989) *Nature*, **339**, 483–484.