Internal Mobility in the Partially Folded DNA Binding and Dimerization Domains of GAL4: NMR Analysis of the N−H Spectral Density Functions[†]

J.-F. Lefèvre,[‡] K. T. Dayie,[§,||] J. W. Peng,[⊥] and G. Wagner*[,§,||]

Committee on Higher Degrees in Biophysics, Harvard University, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, Vertex Pharmaceuticals Inc., 40 Allston Street, Cambridge, Massachusetts 02139, and Groupe de Cancérogenèse et de Mutagenèse Moléculaire et Structurale, ESBS, Bd Sébastien Brant, 67400 Strasbourg-Illkirch Cedex, France

Received November 9, 1995; Revised Manuscript Received January 4, 1996[⊗]

ABSTRACT: The DNA binding domain (residues 1−65) of the yeast transcriptional activator GAL4 is only partially folded. While residues 10−41, the DNA recognition domain, form a well-defined structure in the free protein, the whole polypeptide folds up and dimerizes upon binding DNA. In order to describe the mobility of the protein, we have characterized the frequency spectrum of the motions of N−H bond vectors of GAL4(1−65) using a reduced spectral density mapping approach (an approximation of the full spectral density mapping technique) [Peng, J. W., & Wagner, G. (1992a) *J. Magn. Reson. 98*, 308−332; Peng, J. W., & Wagner, G. (1992b) *Biochemistry 31*, 8571−8586]. ^{15}N spin−lattice relaxation [$R_N(N_z)$], spin−spin relaxation [$R_N(N_{x,y})$], cross-relaxation [$R_N(H_z \rightarrow N_z)$], two-spin order [$R_{NH}(2H_zN_z)$], and antiphase [$R_{NH}(2H_zN_{x,y})$] rates were determined for 52 of the 65 backbone amide groups at 10 °C and pH 6.5 at 11.74 T. Calculations of the spectral density functions using a reduced set of $R_N(N_z)$, $R_N(N_{x,y})$, $R_N(H_z \rightarrow N_z)$, and $R_{NH}(2H_zN_z)$ gave excellent agreement with those calculated using all six sets. The reduced method has the added advantage that the errant behavior seen at high field values is circumvented. A linear correlation was found between $J(\omega_N)$ and $J(0)$ with a limited and clearly defined range of $J(0)$ values which defines the range of rates for internal motions in GAL4(1−65). It appears that all residues experience a combination of two movements: one of the overall tumbling (correlation time, 8.65 ns) and the other of fast internal fluctuations of the structure. The respective weights of these contributions vary with the primary sequence and faithfully mirror the secondary and tertiary elements of the protein. The position on the correlation line of $J(\omega_N)$ versus $J(0)$ indicates the amount of angular averaging relative to the overall motion of the protein. A spectral density function for internal motions can be described.

Increasingly, two-dimensional (2D)[1] nuclear magnetic resonance (NMR) is being used to study the dynamics of proteins (Wagner, 1993). Among the motivations for these studies are to search for correlations between structure, dynamics, and function [e.g. Karplus and McCammon (1983) and Frauenfelder et al. (1988)] and to investigate the effect of internal motion on NMR structure calculations. NMR relaxation is ideally suited to probe, on a residue by residue basis, the vast range of molecular motions spanning picoseconds to hours [see for example Tycko (1994)]. Until recently, these motions were probed by measuring the three relaxation parameters T_1, T_2, and NOE for the backbone ^{15}N and ^{13}C nuclei; then the experimentally measured relaxation times were used to fit the parameters of a model. Among them were the "wobbling in a cone" model (Kinoshita et al., 1977; Richarz et al., 1980) or the so-called "model free approach" of Lipari and Szabo (1982a,b). Both approaches make an a priori assumption about a simple functional form of the spectral density function, $J(\omega)$, depending on some parameters, such as an overall rotational correlation time, τ_c, an order parameter, S^2, a cone opening angle, etc. Recently, Peng and Wagner (1992a,b) have described a more direct approach in which the spectral density functions are evaluated at select frequencies independent of a motional model or an analytical form of the spectral density function. In this approach, it was proposed to measure more relaxation parameters in addition to T_2s, T_1s, and NOEs. Originally, it was proposed to measure in addition the relaxation rate of longitudinal two-spin order, transverse relaxation rates of antiphase coherence, and the proton spin−lattice relaxation rate. This approach was called the spectral density mapping technique. Here, we propose a slightly modified version, reduced spectral density mapping which allows the calculation of the spectral density function based on any three or four members of the set of six relaxation rates (vide infra). This approach is operationally identical to quasi spectral density mapping proposed recently by Ishima and Nagayama (1995) and the high-frequency approximation of Farrow et al. (1995), who propose using only three rates; ours (using the four rates) allows the estimate of the longitudinal relaxation rate of the proton, a measurement which is otherwise fraught with errors.

[†] This work was supported by grants to G.W. from NSF (Grant MCB9316938) and NIH (Grant GM47467) and to J.E.L. by UICC.

* Author to whom correspondence should be addressed.
[‡] ESBS.
[§] Harvard Medical School.
[||] Harvard University.
[⊥] Vertex Pharmaceuticals Inc.

[⊗] Abstract published in *Advance ACS Abstracts,* February 15, 1996.

[1] Abbreviations: 2D, two-dimensional; CPMG, Carr−Purcell−Meiboom−Gill; CSA, chemical shielding anisotropy; DD, dipole−dipole; FID, free induction decay; GARP, globally optimized alternating phase rectangular pulse; INEPT, incredible enhancement by polarization transfer; NOE, nuclear Overhauser effect; TPPI, time-proportional phase incrementation.

0006-2960/96/0435-2674$12.00/0

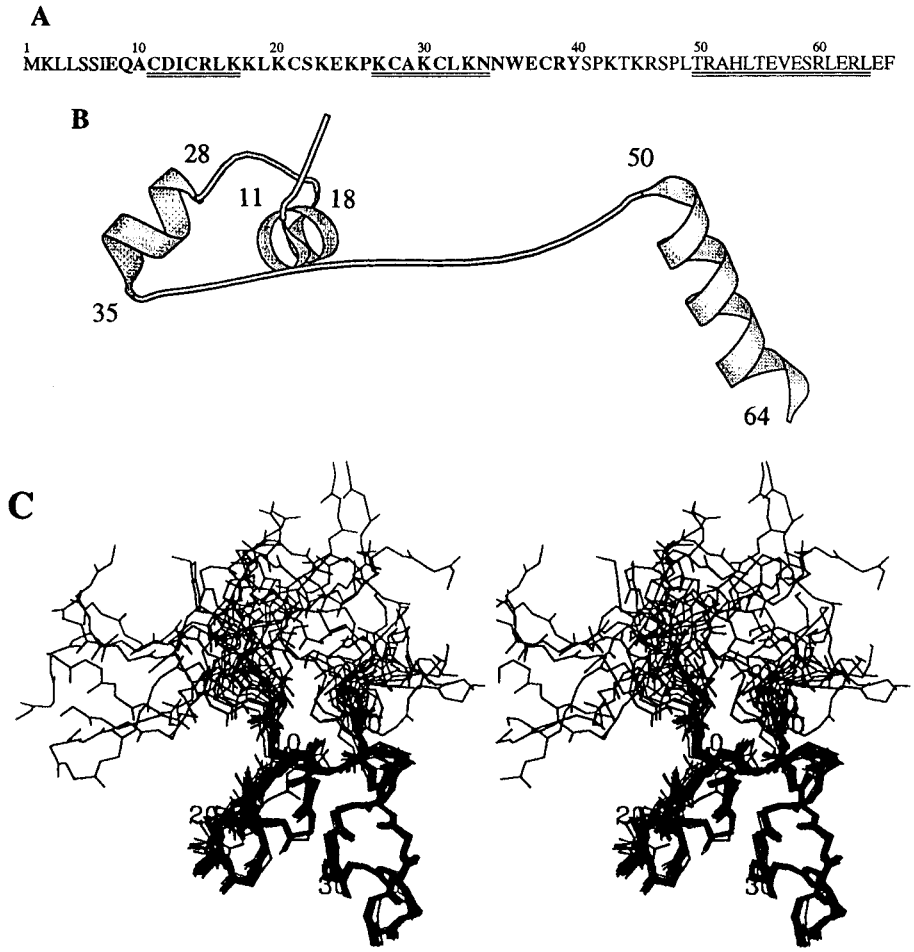

FIGURE 1: (A) Amino acid sequence of GAL4(1−65). Phe66 is a cloning artifact and not part of the wild-type sequence. Residues in bold represent the part of the protein that is folded in free GAL4(1−65) in solution. Sequences underlined represent the three helices observed in the crystal structure of the GAL4(1−65) dimer bound to DNA (Marmorstein et al., 1992). (B) Ribbon diagram of GAL4(1−65) in the crystal structure of the DNA complex. The residues at the ends of the helices are identified with sequence numbers. (C) Stereoview of an ensemble of 22 NMR models of GAL4(1−65) representing the solution structure of the free protein. The backbone of residues 1−43 is shown. The well-defined part extends from Q9 to Y40.

Here, we have applied this technique to the N-terminal DNA binding and dimerization domain (residues 1−65) of the transcriptional activator GAL4 which regulates galactose catabolism in the yeast *Saccharomyces cerevisiae* (Bram & Kornberg, 1985; Giniger et al., 1985; Johnston, 1987a; Carey et al., 1989) and carries out several functions associated with the different regions of the polypeptide chain with the DNA binding localized in residues 1−65 (Kegan et al., 1986; Johnston, 1987b). The N-terminal fragment under study (residues 1−65) is monomeric in solution but binds as a dimer to each 17-base pair (bp) palindromic consensus sequence of the eleven known GAL4 binding sites (Giniger et al., 1985; Bram & Kornberg, 1985; Fedor et al., 1988).

Free GAL4(1−65) in solution is only partially folded. The three-dimensional solution structure of the free amino terminal (1−65) fragment as determined by NMR (Baleja et al., 1992) indicates that residues 9−40 form a compact, globular metal-binding domain while the first nine residues and the residues C-terminal to Ser41 (42−65) are disordered (Figure 1C). On the other hand, the whole polypeptide folds up and dimerizes when binding DNA (Figure 1B). The amino acid sequence of Gal4(1−65) is shown in Figure 1A. Residues that form α-helices in the crystal structure of GAL4(1−65) in the complex with DNA (Marmorstein et al., 1992) are underlined. The fragment that is folded in the free protein, we call it DNA recognition domain, consists of two short helices from Asp12−Leu19 and Lys30−Trp36 (Baleja et al., 1992; Kraulis et al., 1992; Shirakawa et al., 1993). They pack against two central zinc ions to form a bimetal−thiolate cluster using the six conserved cysteines. Each of the stretches of residues 10−22 and 27−39 forms a similar α-helix/tight turn/extended chain structure. As a result, their backbone atoms can be superimposable to within 0.89 Å, suggesting an internal 2-fold symmetry. The first eight residues and residues C-terminal to Lys43 have very poor chemical shift dispersion in the homonuclear 2D data sets, show very little long range NOEs to other parts of the protein, and have chemical shifts close to random coil values (K. T. Dayie and G. Wagner, unpublished results).

The studies described here focus on the relative mobility within free GAL4(1−65) using ^{15}N relaxation experiments. We were motivated to investigate this apparent disorder seen in the free solution form with a view to establishing whether this disorder is due to greater mobility of these regions or simply due to a lack of experimental constraints or both. We have measured all six relaxation parameters to analyze the frequency spectrum of internal dynamics in the DNA binding domain of this GAL4 fragment. We found that the power spectral density functions at different frequencies are correlated and the positions on the correlation diagram are

characteristic for the degree of angular averaging relative to the overall motion of the protein. We find a well-defined range of $J(0)$ values which defines the range of time constants of internal motions in GAL4(1−65).

EXPERIMENTAL PROCEDURES

Protein Sample Preparation. The DNA binding domain of GAL4 was overexpressed in the *Escherichia coli* strain XA-90 from a plasmid vector carrying the gene conferring ampicillin resistance and the coding region of the N-terminal 65 residues of GAL4 under the control of the *tac* promoter. This construct carries the first 65 residues of GAL4 and an additional Phe (F) introduced as a 66th residue by a cloning artifact [originally described by Carey et al. (1989)]. Experiments were performed on a 1.5 mM sample at pH 6.5.

NMR Spectroscopy. All NMR spectra were recorded at 283 K on a Bruker AMX 500 spectrometer. Following earlier protocols (Peng & Wagner, 1992a,b), the following six relaxation parameters were measured for the backbone N−H vectors of the GAL4 DNA binding protein. The longitudinal ^{15}N relaxation rate $R_N(N_z)$ was obtained with eleven parametric delays of 20 (2×), 100, 120, 200, 300, 400, 500, 700, 800, and 3000 ms. For the in-phase ^{15}N transverse relaxation rate $R_N(N_{x,y})$, nine delays of 4, 8, 20, 32, 48, 80, 112, 144, and 200 ms were used. The heteronuclear cross-relaxation rate constants $R_N(H_z \rightarrow N_z)$ (^{1}H−^{15}N heteronuclear NOE) were obtained with two spectra recorded with 3 s of proton presaturation and two others without presaturation. The antiphase ^{15}N transverse relaxation rate $R_{NH}(2H_zN_{x,y})$ was obtained with ten delays of 2, 10, 20, 40 (2×), 60, 80, 120, 160, and 200 ms. For the decay rate of the longitudinal two-spin order $R_{NH}(2H_zN_z)$, delays of 2, 10, 20, 40 (2×), 60, 80, 120, 160, and 200 ms were used. For the longitudinal amide proton spin−lattice relaxation rate $R_H(H_z)$, thirteen delays of 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, and 150 ms were used.

In general, the relaxation rates were measured in a series of heteronuclear ^{1}H−^{15}N correlated spectra using an INEPT at the beginning and a reverse INEPT at the end of the sequence to enhance sensitivity. Details of the pulse sequences have been described previously (Peng & Wagner, 1992a,b). Each 2D data set consists of 2048 real data points in t_2 for 128 t_1 blocks with the ^{15}N sweep width set to 1500 Hz. Thirty-two transients per t_1 increment were recorded with 64 dummy scans acquired prior to the start of each experiment for all the measurements except for the $\{^1H\}^{15}N$ NOE experiments which used 128 scans per t_1 increment. Quadrature phase detection in the ^{15}N dimension was achieved by the time-proportional phase incrementation method of Marion and Wüthrich (1983). The total acquisition time for the longitudinal relaxation rate ^{15}N and NOE experiments was typically 2−4 days and 1 day for the other relaxation series. All 2D data sets were processed on both SUN SLC and IRIS 4D-35 workstations using the program Felix 2.05 provided by Dennis Hare (Hare Research, Seattle). Sixty degree-shifted sine-squared window functions were applied to each free induction decay (FID), zero-filled to 8192 real points prior to Fourier transformation, and then strip transformed in the recorded dimension.

Cross-peak intensities of the 2D spectra were determined from peak heights integrated along the ^{1}H(F_2) dimension. The uncertainties, due to random errors, in the measured heights were determined from Monte Carlo simulations. The relaxation rate constants were obtained from nonlinear fits to monoexponential functions using the Levenburg−Marquardt algorithm (Press et al., 1988) as described previously (Peng & Wagner, 1992b).

RESULTS AND DISCUSSION

Use of a Reduced Spectral Density Function Defined at 0, ^{15}N, and ^{1}H Frequencies. The relaxation mechanism for NH spin systems arises from the local time-varying magnetic fields generated at the ^{1}H and ^{15}N nuclei due to global tumbling and internal mobility of the various N−H bond vectors. The frequency spectrum of the rotational motions of the N−H vector in a single molecule relative to the external superconducting magnetic field is given by the Fourier transform of the spherical harmonics describing the motion, the spectral density functions. This spectral density function describes how the dipole−dipole and CSA energy of interactions between N and H is distributed at the five different angular frequencies: 0, ω_N, $\omega_H + \omega_N$, ω_H, and $\omega_H - \omega_N$. As has been shown earlier (Peng & Wagner, 1992a,b), the set of six relaxation rates $R_N(N_z)$, $R_N(N_{x,y})$, $R_N(H_z \rightarrow N_z)$, $R_{NH}(2N_zH_z)$, $R_{NH}(2N_{x,y}H_z)$, and $R_H(H_z)$ allows the unambiguous calculation of the spectral density functions $J(\omega)$ via the equation (Peng & Wagner, 1992a)

$$\begin{pmatrix} R_N(N_z) \\ R_N(N_{x,y}) \\ R_N(H_z \rightarrow N_z) \\ R_{NH}(2N_zH_z) \\ R_{NH}(2N_{x,y}H_z) \\ R_H(H_z) \end{pmatrix} = \begin{pmatrix} 0 & A & E & 0 & 6A & 0 \\ \frac{2E}{3} & \frac{A}{2} & \frac{E}{2} & 3A & 3A & 0 \\ \frac{2E}{3} & \frac{A}{2} & \frac{E}{2} & 0 & 3A & 1 \\ 0 & 0 & E & 3A & 0 & 1 \\ 0 & A & 0 & 3A & 6A & 1 \\ 0 & -A & 0 & 0 & 6A & 0 \end{pmatrix} \times \begin{pmatrix} J(0) \\ J(\omega_N) \\ J(\omega_{H^N} - \omega_N) \\ J(\omega_{H^N}) \\ J(\omega_{H^N} + \omega_N) \\ \rho_{H^NH^i} \end{pmatrix} \quad (1)$$

where $A = \gamma_{H^N}{}^2\gamma_N{}^2\hbar^2/4r_{NH^N}{}^6$, $B = \Delta^2\omega_N{}^2/3$, and $E = 3A + B$. For a CSA value of $\Delta = -160$ ppm (Hiyama et al., 1988) and in internuclear ^{15}N−^{1}H bond distance of $r_{NH^N} = 1.02$ Å (Keiter, 1986), the constants A and B become $\approx 1.3 \times 10^9$ and 0.9×10^9 (rad/s)^2, respectively, at 500 MHz. The right-hand column vector consists of the unknown $J(\omega)$ samplings that can be solved for.

Equation 1 represents the situation where no chemical exchange contributes to the transverse nitrogen relaxation rates. Depending on the way the transverse relaxation rates $R_N(N_{x,y})$ and $R_{NH}(2N_{x,y}H_z)$ are measured, slow exchange between different states may artificially increase these rates if, for example, a CPMG-type experiment is run without limiting the interpulse delays to $k\delta \ll 1$, where k is the exchange or scalar relaxation rate constant and δ is CPMG spin echo time (Bloom et al., 1965). To avoid these complications, we measured these rates with spin locks of the nitrogen coherence. If a sufficiently high spin lock power

is available and the nitrogen carrier is placed on resonance, these artifactual exchange contributions to the transverse relaxation rates can essentially be eliminated. Following the procedures of Deverell et al. (1970), exchange contributions to the transverse relaxation rates can be identified with measurements at variable spin lock field strengths (K. T. Kayie and G. Wagner, in preparation). From the extrapolation to infinite B_1 field strength, the transverse relaxation rates can be separated from exchange contributions which are therefore not considered in the following treatment.

A characteristic feature of the spectral density function is that the integral over the whole frequency range is a constant. This follows from the conservation of the total mean dipole−dipole and CSA energy available to an ensemble of independent molecules. As a result, the density of the frequency of motions is shifted toward lower frequencies for slowly tumbling molecules, i.e. larger-sized macromolecules. As has been confirmed experimentally by the recent calculation of the spectral density function observed for N−H vectors in eglin C (Peng & Wagner, 1992a,b), at high frequencies (around 500 MHz), the spectral density function is theoretically quite flat. This slow variation of $J(\omega)$ around the proton frequency (at 300, 500, and 600 MHz) suggests that it might be possible to equate the three values of the spectral density function at $\omega_H - \omega_N$, ω_H, and $\omega_H + \omega_N$ into an average value, expressed hereafter by $\langle J(\omega_H)\rangle$. Thus, three different relaxation rates are sufficient to determine the three values of the spectral density functions $J(0)$, $J(\omega_N)$, and $\langle J(\omega_H)\rangle$. Alternatively, more relaxation parameters could be measured, and the values of the spectral density function at the three frequencies could be obtained from a least squares fit.

Reduced Spectral Density Function from Less Than Six Relaxation Parameters. When the reduced spectral density function is to be determined from three relaxation rates, one has the obvious choice of $R_N(N_z)$, $R_N(N_{x,y})$, and $R_N(H_z \rightarrow N_z)$. Such a combination of rates has the advantage that it is independent of proton relaxation due to dipolar interaction with other protons.

If $\mathbf{J}_{red,3}$ is made the column vector containing the three values of $J(\omega)$ and $\mathbf{R}_{red,3}$ the vector containing the classically measured $R_N(N_z)$, $R_N(N_{x,y})$, and $R_N(H_z \rightarrow N_z)$, the matrix of coefficients, $\mathbf{C}_{red,3}$, relates the two via the equation

$$\mathbf{R}_{red,3} = \mathbf{C}_{red,3} \times \mathbf{J}_{red,3} \tag{2}$$

or explicitly

$$\begin{pmatrix} R_N(N_z) \\ R_N(N_{x,y}) \\ R_N(H_z \rightarrow N_z) \end{pmatrix} = \begin{pmatrix} 0 & E & 7A \\ \frac{2E}{3} & \frac{E}{2} & \frac{13A}{2} \\ 0 & 0 & 5A \end{pmatrix} \begin{pmatrix} J(0) \\ J(\omega_N) \\ \langle J(\omega_H)\rangle \end{pmatrix} \tag{2a}$$

where $\mathbf{C}_{red,3}$, the matrix of coefficients, is calculated by replacing each of the spectral density values $J(\omega_N+\omega_H)$, $J(\omega_H)$, and $J(\omega_N-\omega_H)$ with the averaged value $\langle J(\omega_H)\rangle$ and $\mathbf{J}_{red,3}$ is calculated by simply inverting eq 2:

$$\mathbf{J}_{red,3} = \mathbf{C}_{red,3}^{-1} \times \mathbf{R}_{red,3} \tag{3}$$

with

$$\mathbf{C}_{red,3} = \begin{pmatrix} -\frac{3}{4E} & \frac{3}{2E} & -\frac{9}{10E} \\ \frac{1}{E} & 0 & -\frac{7}{5E} \\ 0 & 0 & \frac{1}{5A} \end{pmatrix} \tag{3a}$$

Additionally, one of the rates, $R_{NH}(2N_{x,y}H_z)$, $R_{NH}(2N_zH_z)$, or $R_H(H_z)$, could be used. Any of these rates is dependent on the relaxation by dipolar interaction between non-amide protons and protons directly attached to the nitrogen (Peng & Wagner, 1992a,b). We call the relaxation vector that contains the additional relaxation rate $\mathbf{R}_{red,4}$. Analogous to eq 3, the spectral density function is given by

$$\mathbf{J}_{red,4} = \mathbf{C}_{red,4}^{-1} \times \mathbf{R}_{red,4} \tag{4}$$

$\mathbf{J}_{red,4}$ now contains the same terms as $\mathbf{J}_{red,3}$ with an additional fourth term ρ which is the dipolar interaction with other protons contributing to the relaxation of the amide proton (Peng & Wagner, 1992). The coefficient matrix is similar to $\mathbf{C}_{red,3}$ with an extra row and column. For example, if $R_{NH}(2N_zH_z)$ is added to the three previous relaxation rate constants, we have

$$\mathbf{C}_{red,4} = \begin{pmatrix} 0 & E & 7A & 0 \\ \frac{2E}{3} & \frac{E}{2} & \frac{13A}{2} & 0 \\ 0 & 0 & 5A & 0 \\ 0 & E & 3A & 1 \end{pmatrix} \tag{5}$$

$$\mathbf{C}_{red,4}^{-1} = \begin{pmatrix} -\frac{3}{4E} & \frac{3}{2E} & -\frac{9}{10E} & 0 \\ \frac{1}{E} & 0 & -\frac{7}{5E} & 0 \\ 0 & 0 & \frac{1}{5A} & 0 \\ -1 & 0 & \frac{4}{5} & 1 \end{pmatrix} \tag{5a}$$

Alternatively, all five nitrogen relaxation rate constants can be used in a least squares fitting procedure. If $\mathbf{R}$ is the column vector containing the five nitrogen relaxation rate constants, in the same order as in eq 1 without the proton relaxation rate $R_H(H_z)$, the spectral density function is given by

$$\mathbf{J}_{red,5} = ({}^t\mathbf{C} \times \mathbf{C})^{-1} \times ({}^t\mathbf{C} \times \mathbf{R}) \tag{6}$$

where

$$\mathbf{C} = \begin{pmatrix} 0 & E & 7A & 0 \\ \frac{2E}{3} & \frac{E}{2} & \frac{13A}{2} & 0 \\ 0 & 0 & 5A & 0 \\ 0 & E & 3A & 1 \\ \frac{2E}{3} & \frac{E}{2} & \frac{7A}{2} & 1 \end{pmatrix} \tag{7}$$

and ${}^t\mathbf{C}$ is the transpose of $\mathbf{C}$.

Test of the Reduced Spectral Density Approach with Synthetic, Randomly Shaped Spectral Density Functions. The use of the reduced spectral density function was tested by construction of various spectral density functions randomly

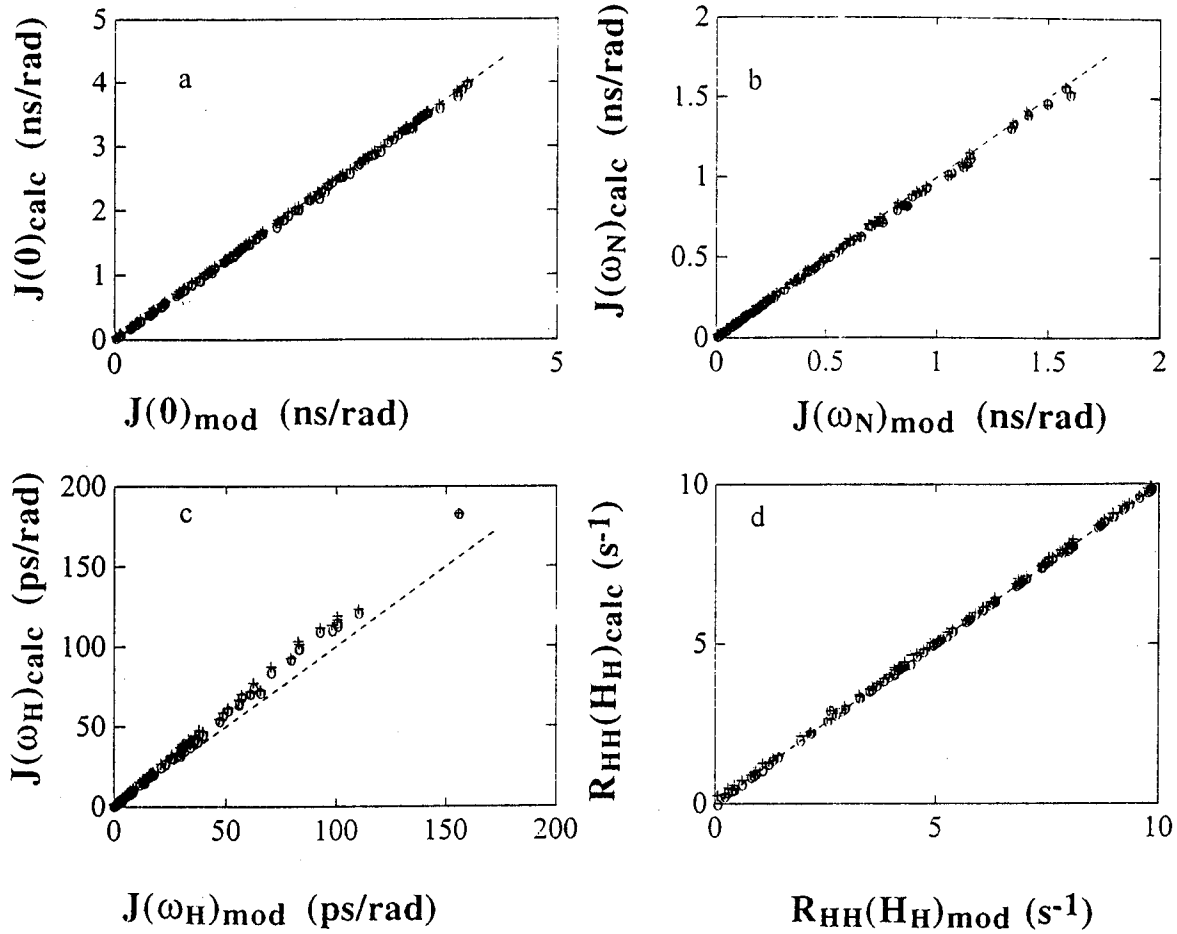

FIGURE 2: Plots of calculated versus model values of randomly shaped (see text) spectral density functions. Open circles are the values calculated with the direct method (eqs 3 and 6), and crosses represent values obtained with the least squares method (eq 8). Plots are shown for the spectral density function values at (a) 0, (b) ω_N, and (c) $\langle\omega_H\rangle$ and (d) for the proton longitudinal relaxation rate ρ. The dashed line represents a line with slope 1.

shaped, with the only constraint that the function should decrease monotonically, subject to the following conditions:

$$0 < J(0) < 4 \times 10^{-9}$$

$$0 < J(\omega_N) < J(0)/2$$

$$0 < J(\omega_H+\omega_N) < J(\omega_N)/10 \qquad (8)$$

$$0 < J(\omega_H) < J(\omega_H+\omega_N)$$

$$0 < J(\omega_H-\omega_N) < J(\omega_H)$$

$R_N(N_z)$, $R_N(N_{x,y})$, $R_N(H_z \rightarrow N_z)$, $R_{NH}(2N_zH_z)$, and $R_{NH}(2N_{x,y}H_z)$ were then calculated from these $J(\omega)$ values, and the vectors $\mathbf{J}_{red,3}$, $\mathbf{J}_{red,4}$, and $\mathbf{J}_{red,5}$ were determined using eqs 3, 4, and 6. The results obtained on 100 randomly chosen functions are shown in Figure 2a−d. The correlations between the recalculated and the model values of $J(0)$ and $J(\omega_N)$ are excellent, no matter what method is used. The values resulting from the direct calculation of $\mathbf{J}_{red,3}$ and $\mathbf{J}_{red,4}$ are identical. The values of $\langle J(\omega_H)\rangle$ which correlate with the highest value of the spectral density function in the high-frequency range $J(\omega_H+\omega_N)$ are slightly overestimated by the method. This is expected if the spectral density function is not entirely flat but decreases in the high-frequency range (Peng & Wagner, 1995). Again, the values extracted from $\mathbf{J}_{red,3}$ and $\mathbf{J}_{red,4}$ are totally identical. It should be noticed that the value $\langle J(\omega_H)\rangle$ only depends on the cross-relaxation rate constants [$R_N(H_z \rightarrow N_z)$], as can be seen from eqs 5 and 7. This dependency has been reported independently by Farrow et al. (1995b). The good correlation that we observe for $\langle J(\omega_H)\rangle$ then becomes obvious, as $J(\omega_H+\omega_N)$ is dominant in the cross-relaxation rate constant. This probably explains why we do not observe an obvious correlation between $\langle J(\omega_H)\rangle$ and $J(\omega_H)$ or $J(\omega_H-\omega_N)$ (data not shown). Note that $\langle J_{avg}(\omega_H)\rangle$ used in this work corresponds to the $J(\omega_H+\omega_N)$ of Ishima and Nagayama (1995a) and $J(0.87\omega_H)$ of Farrow et al. (1995b). Also, the assumption that $J(\omega) \propto \omega^{-2}$ for the high-frequency components (Farrow et al., 1995b) is not necessary if one uses $J(\omega_H+\omega_N)$ for the spectral density function at high frequencies. Calculations using either approach yield identical results (data not shown).

The behavior of the various calculations with respect to the corruption of the data by noise was tested using a Monte Carlo simulation. Some of the spectral density functions constructed as described above were chosen, covering a wide range of $J(\omega)$ values. Random Gaussian noise was added to the relaxation rate constants calculated from each spectral density function. The maximum amplitude of the noise was set to 5% for all relaxation rate constants except for the proton relaxation rate (10%) and the cross-relaxation rate which is calculated from the nuclear Overhauser effect and which was therefore assigned a maximum noise of 20%. A total of 100 noisy samples of the same data set were

produced. The average values of the recalculated $J(\omega)$ and the standard deviation were determined. Typical results are given for one set of data in Table 1 of the supporting information.

The calculation using a reduced set of relaxation rate constants ($\mathbf{R}_{red,4}$) which excludes the cross-relaxation rate constant $R_N(H_z \rightarrow N_z)$ gives very poor results. This is because, in this case only, the value of $\langle J(\omega_H)\rangle$ which is very small is determined from a combination of large relaxation rate constant values that are comparable in magnitude. The least squares method also does not perform very well in this statistical test, giving fairly large standard deviations. One of the best results was given by the combination of $R_N(N_z)$, $R_N(N_{x,y})$, $R_N(H_z \rightarrow N_z)$, and $R_{NH}(2N_zH_z)$. The coefficient matrix for this calculation is given in eq 5. In the following, we use this set of four experimental relaxation rate constants to determine the spectral density functions for the N−H vectors in GAL4(1−65).

Theoretical Features of the Spectral Density Functions. The sets of calculated $J(\omega)$ values are informative of the shape of the spectral density functions. In order to analyze these functions in terms of motion of the N−H vectors, we consider the following points.

(i) In flexible molecules such as proteins, the correlation function describing the motion of a vector is expected to be a composite function (Brooks et al., 1987). Internal motion may increase the decay rate of the correlation function, but this motion is generally restricted and does not randomize completely the position of the vector. The overall tumbling of the molecule ultimately brings the correlation function to zero. As a consequence, the spectral density function is itself a composite function. If the various motions are not correlated, one may express $J(\omega)$ as a weighted sum:

$$J(\omega) = \sum_i a_i J_i(\omega) \tag{9}$$

where a_i is the scaling factor of the spectral density functions $J_i(\omega)$ characterizing each motion and $\sum_i a_i = 1$. Several models of $J(\omega)$, all based on eq 9, have been proposed (Woessner, 1962; London & Avitable, 1978; Tropp, 1980; Jardetzky, 1981; Lipari & Szabo, 1982a,b) which have various definitions of a_i and $J_i(\omega)$.

(ii) The value $J(0)$ is equal to the area under the correlation function $G(t)$. For an isotropic Brownian motion, the correlation function is a simple decaying exponential (Abragam, 1961):

$$G(t) = {}^1/_5 \exp\left(-\frac{t}{\tau_c}\right) \tag{10}$$

with τ_c being the correlation time. The corresponding spectral density function is

$$J(\omega) = \frac{2}{5}\frac{\tau_c}{1 + \omega^2\tau_c{}^2} \tag{11}$$

and its value at zero frequency is

$$J(0) = 2\int_0^\infty G(t)\,dt = {}^2/_5\tau_c \tag{12}$$

If internal motion occurs on a time scale smaller than the overall tumbling of the molecule which is characterized by τ_c, $G(t)$ decreases more rapidly than the exponential given in eq 10. Thus, the area under the correlation function and the $J(0)$ value decrease (see eq 12). Because of fast internal motions in proteins, $J(0)$ will always be lower than $2\tau_c/5$, and the range of $J(0)$ values has an upper limit in a protein, given by the overall correlation time.

(iii) The area under the spectral density function $J(\omega)$ is constant and equal to $G(0)$:

$$\int_{-\infty}^{\infty} J(\omega)\,d\omega = G(0) = {}^1/_5 \tag{13}$$

As a consequence, a decrease of the $J(0)$ value is compensated for by an increase of the value of $J(\omega)$ at high frequencies. The function is intense in the low frequencies for slow overall tumbling and flattens when the tumbling rate increases.

(iv) Due to the nature of our calculation, the value obtained for the spectral density function at zero frequency may contain an adiabatic contribution from motions which are slower than the overall tumbling of the molecule and lie in the microsecond to millisecond range. In the presence of fast intermediate chemical exchange, the transverse relaxation rate is increased by this exchange term, R_{ex}, which modifies the calculated value of spectral density function $J'(0)$; thus

$$J'(0) = J(0) + \frac{2R_{ex}}{3E} \tag{14}$$

E is defined in eq 1. To eliminate exchange contributions, transverse relaxation rates can be conveniently recorded under spin lock conditions. The strength of the spin lock field, ω_1, determines to which extent the exchange contributions can be eliminated. For an exchange between two states, A and B, characterized by the on and off rates k_1 and k_{-1}, respectively, a resonance frequency difference $\Delta\nu$, and populations P_A and P_B in the A and B states, the expression of R_{ex} is (Sandström, 1982)

$$R_{ex} = \frac{4\pi^2\Delta\nu^2 P_A P_B \tau_{ex}}{1 + (\tau_{ex}\omega_1)^2} \tag{15}$$

where $(k_1 + k_{-1})^{-1} = \tau_{ex}$. This equation is valid for a spin lock field applied on resonance. Alternatively, R_{ex} can be estimated from a multiple spectrometer study as was recently discussed in detail for eglin C [see eq 30 in Peng and Wagner (1995)].

(v) In the analysis of the experimental values of the spectral density function, we will assume two main features. First, the spectral density function for the overall motion is well-represented by a Lorentzian, which means that the correlation function can be approximated with a monoexponential decay. This is not a severe assumption as the shape of the protein is not far from being globular. Second, internal motions are expected to be fast such that the corresponding spectral density function is flat in the low-frequency range. Hence, in the frequency range up to the nitrogen frequency (50 MHz spectrometer nitrogen frequency), it can be represented by a Lorentzian which is a flat function for small correlation times. However, we will make no other assumption on the shape of the spectral density function for internal dynamics. We will then decompose the spectral density

function in two parts: one for the overall tumbling $J_0(\omega)$ and the other for internal motion made of several contributions $J_i(\omega)$:

$$J(\omega) = a_0 J_0(\omega) + \sum_j a_j J_j(\omega) \qquad (16)$$

Measurements of the Relaxation Rate Constants and Determination of the Corresponding Reduced Spectral Density Functions. As mentioned earlier, the solution structure of GAL4(1−65) (Baleja et al., 1992) indicates a well-defined core which contains two helices separated by a large loop. These secondary structure elements are held together by a cluster of two zinc atoms chelated by six cysteines. On each side of the core, the structured polypeptide chain extends only a few amino acids, with the result that the first 8 N-terminal and last 23 C-terminal residues are unstructured. Figure 3a−c shows the six relaxation rate constants [$R_N(N_z)$, $R_N(N_{x,y})$, $R_N(H_z \rightarrow N_z)$, $R_{NH}(2N_zH_z)$, $R_{NH}(2N_{x,y}H_z)$, and $R_H(H_z)$] that were measured for GAL4(1−65). The location of the three helices in the crystal structure of the GAL4(1−65)/DNA complex (Marmorstein et al., 1992) is indicated on top of the figure. The part of the polypeptide that is not folded in free GAL4(1−65) is indicated with broken lines. Figure 3a compares $R_N(N_{x,y})$, $R_N(N_z)$, and $R_N(H_z \rightarrow N_z)$ and illustrates the relative size of these rates. While $R_N(N_{x,y})$ clearly discriminates between the folded and flexible parts of the protein, $R_N(N_z)$ does not. The cross-relaxation rates, $R_N(H_z \rightarrow N_z)$, also discriminate between the structured and flexible parts of the protein. Since this is not visible in the scale of Figure 3a, these rates are plotted separately in Figure 3b. The variation of $R_N(H_z \rightarrow N_z)$ with sequence is opposite to that of $R(N_{x,y})$; flexible parts of the protein have higher-than-average $R_{HN}(H_z \rightarrow N_z)$ rates and lower-than-average $R_N(N_{x,y})$ rates. Figure 3c shows the two-spin rates, $R_{HN}(H_zN_{x,y})$ and $R_{HN}(H_zN_z)$, as well as the longitudinal proton relaxation rate $R_H(H_z)$. For comparison at this scale, $R(N_z)$ is also given here. Clearly, relaxation rates of antiphase coherence are fastest of all measured rates. A complete list of all relaxation rates measured and estimates of the errors are provided in the supporting information.

All five values of the spectral density function at 0, ω_N, $\omega_H + \omega_N$, ω_H, and $\omega_H - \omega_N$ were calculated using eq 6. For many residues, large negative values of $J(\omega)$ at high frequencies were found (data not shown). These aberrant values come from the error propagated in the calculation. The spectral density function around the proton frequency is expected to be small and results, in the calculation, from a combination of large comparable relaxation rates. Small relative errors in the data, therefore, end up as large deviations in this part of the spectral density function. Equating the three high-frequency spectral density functions eliminates this aberrant behavior.

The reduced spectral density functions approach using only four of the measured relaxation rate constants, $R_N(N_z)$, $R_n(N_{x,y})$, $R_N(H_z \rightarrow N_z)$, and $R_{NH}(2N_zH_z)$, was therefore used here. The results shown as a function of the sequence are summarized in Figure 4a−c. As expected, the $J(0)$ and $\langle J(\omega_H)\rangle$ vary largely between the terminal ends and the core of the protein. This is consistent with the primary experimental data, the relaxation rates, since for macromolecules $J(0)$ is the dominant contributor to the transverse relaxation

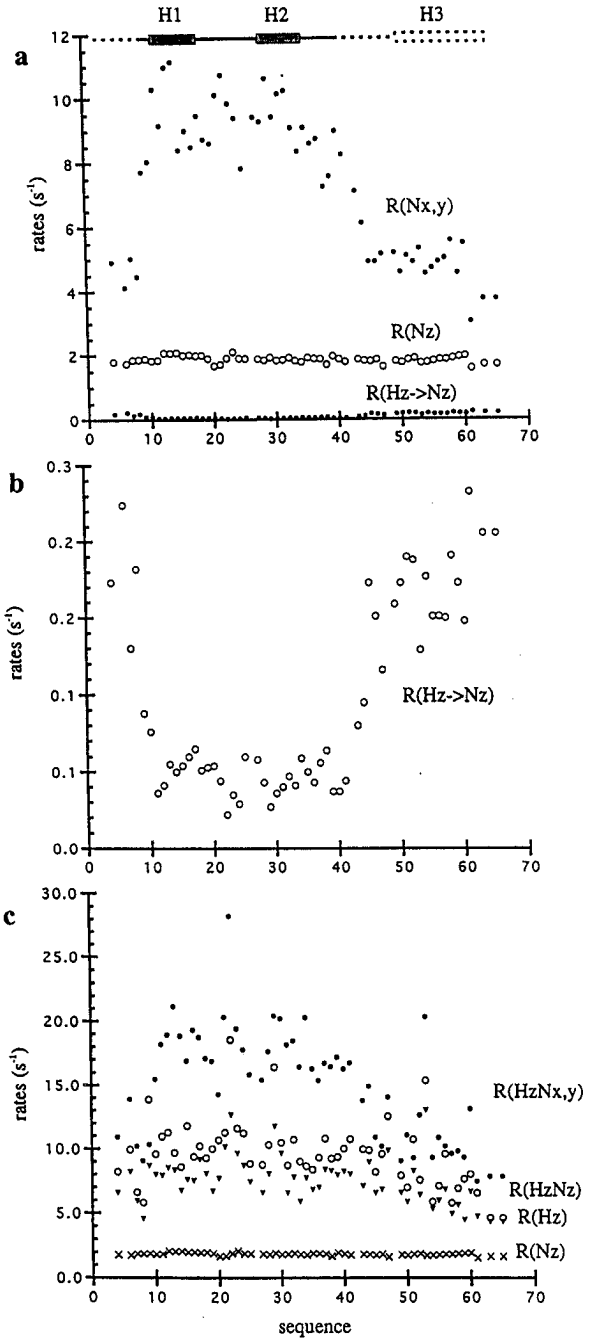

FIGURE 3: Relaxation rates as a function of sequence obtained for GAL4, at 10 °C. (a) Plot at the same scale of $R_N(N_{x,y})$, $R_N(N_z)$, and $R_N(H_z \rightarrow N_z)$. (b) Plot at extended scale of $R_N(H_z \rightarrow N_z)$. (c) Plot of the two-spin relaxation rates $R_{NH}(2N_{x,y}H_z)$ (●) and $R_{NH}(2N_zH_z)$ (○), as well as the longitudinal relaxation rates $R_H(H_z)$ (▼) and $R_N(N_z)$ (×). All rates are in s^{-1}.

rate constant while $\langle J(\omega_H)\rangle$ is directly proportional to the cross-relaxation rate constant, $R_{HN}(H_z \rightarrow N_z)$. $J(\omega_N)$, which dominates the longitudinal ^{15}N relaxation, is quite constant along the protein sequence, as are the longitudinal nitrogen relaxation rates. ρ, the contribution of the dipolar interaction with other protons to the relaxation of the amide proton, exhibits large variations (over 10 s^{-1}) for residues E9, S22, and A29 in the core and S47 and H53 in the C-terminal sequence. The values of the two-spin order relaxation rates, $R_N(2N_zH_z)$, for these five residues exceeded the average for the core residues. This increase in the proton relaxation rate comes probably from the contribution of a fast exchange of the amide proton with the solvent. For example, A29 is the

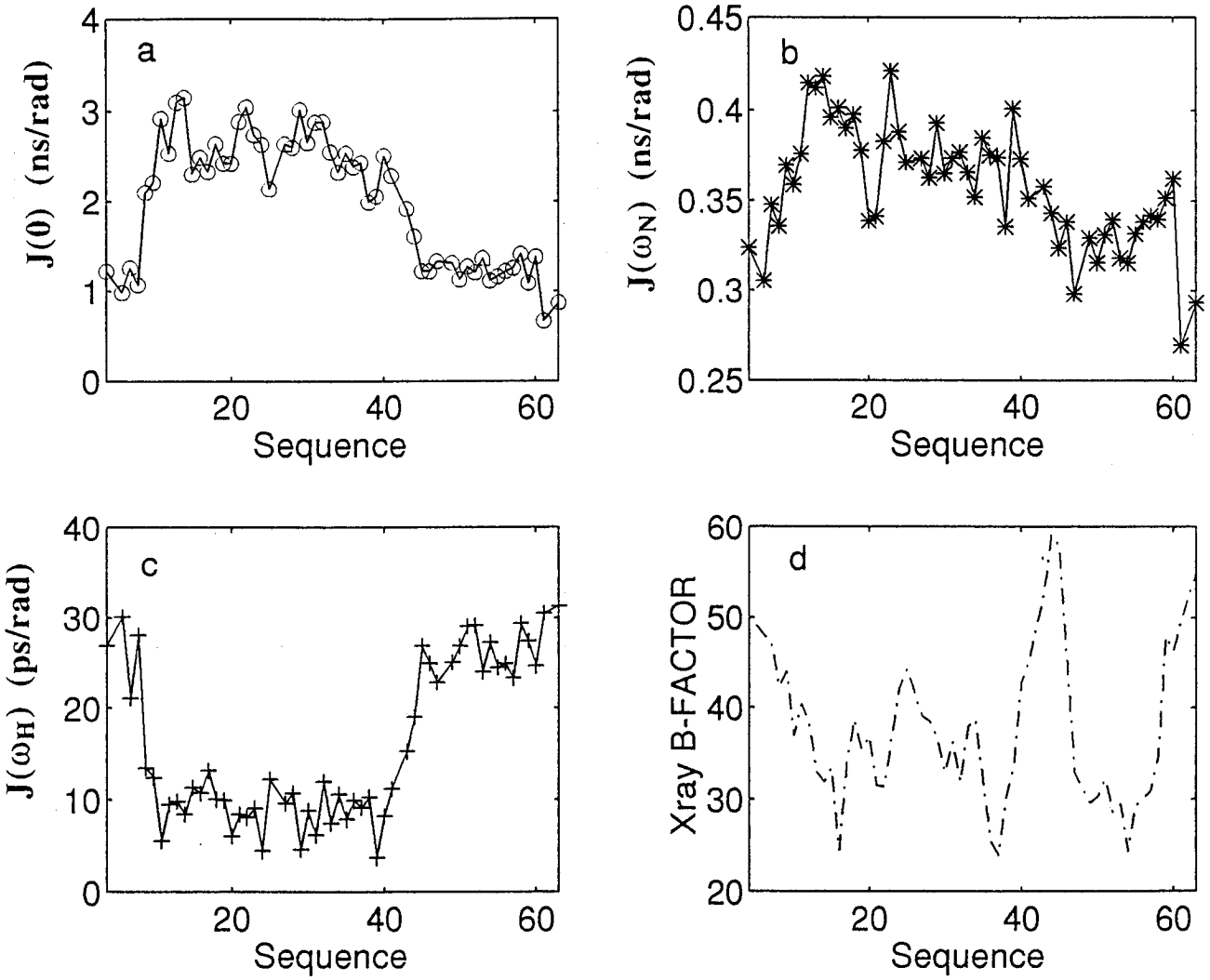

FIGURE 4: Values of the spectral density functions versus the protein sequence at the frequencies of (a) 0, (b) ω_N, and (c) ω_H. For comparison, the *B* factors of the X-ray structure (Marmorstein et al., 1992) are also given (d).

second amino acid of a helix and is not involved in an intramolecular hydrogen bond. Figure 4d shows the *B* factors of the crystal structure of the complex with DNA (Marmorstein et al., 1992). Interestingly, large *B* factors are found at the two termini but also around residues 43; this is the boundary region between the structured and unstructured part of free GAL4(1−65). Helix 3 of the crystal structure has low *B* factors, but the solution structure is clearly flexible in this region.

Analysis of the Spectral Density Functions. To further analyze the measured spectral density functions, the values of $J(\omega_N)$ and $\langle J(\omega_H)\rangle$ are plotted as a function of $J(0)$ in Figure 5. As expected from theoretical considerations, we observe that $J(\omega_N)$ and $\langle J(\omega_H)\rangle$ are linearly correlated to the corresponding $J(0)$ values. The negative slope of $\langle J(\omega_H)\rangle$ as a function of $J(0)$ reflects the compensatory effect between the values of $J(\omega)$ at zero and high frequency (vide supra). $J(0)$ is smaller for the terminal ends than for the core of the protein, while $\langle J(\omega_H)\rangle$ has the opposite behavior. This is indicative of higher mobility in the terminal ends than in the core, in agreement with structure calculations that indicated that the terminal ends (residues 1−8 and 44−63) of the protein appeared "unstructured" (Baleja et al., 1992).

The following analysis shows that, in the low-frequency range, the spectral density functions along the sequence of the protein can be described with a limited number of contributions which are very similar from one residue to the other. Combining the linear equation seen experimentally, we have

$$J(\omega_N) = \alpha J(0) + \beta \tag{17}$$

where α and β are the slope and the intercept of the fit in Figure 5a, respectively. Together with eq 9, we get

$$\sum_i a_i J_i(\omega_N) = \alpha \sum_i a_i J_i(0) + \beta \tag{18}$$

This equation is valid for all residues and consequently for various combinations of a_is (with the condition $\sum_i a_i = 1$). It is easy to show that all the $J_i(\omega)$ components should obey the same linear relationship:

$$J_i(\omega_N) = \alpha J_i(0) + \beta \tag{19}$$

If the $J_i(\omega)$ components are given a Lorentzian shape (eq 11), eq 19 leads to a third degree equation in τ:

$$2\alpha\omega_N{}^2\tau^3 + 5\beta\omega_N{}^2\tau^2 + 2(\alpha - 1)\tau + 5\beta = 0 \tag{20}$$

the roots of which give the correlation time of the various

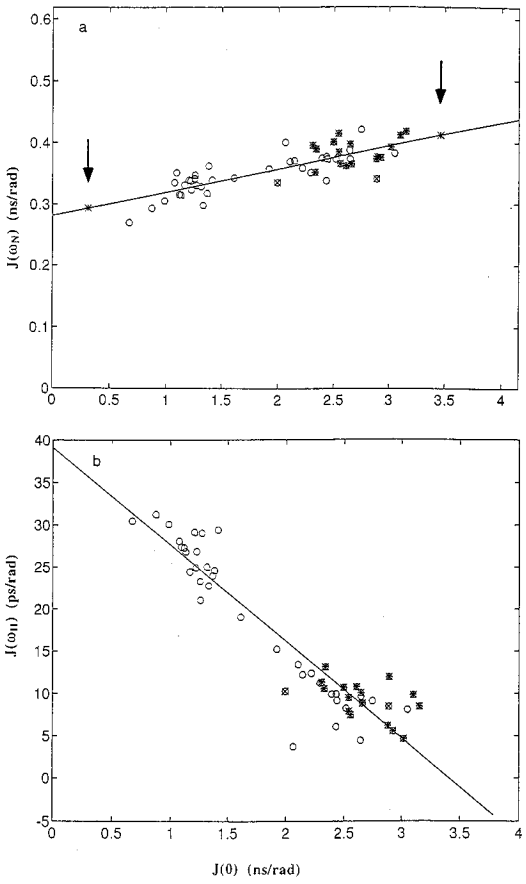

FIGURE 5: Plots of (a) $J(\omega_N)$ and (b) $\langle J(\omega_H)\rangle$ as a function of $J(0)$. The fit was obtained by linear regression. The two stars in a (marked with arrows) indicate points corresponding to the meaningful solution of eq 20. The resulting values of τ_c and τ_i were 8.6 and 0.77 ns, respectively. Residues located in the two helices forming the zinc-binding core of the protein are indicated by the symbol ⊗.

motions contributing to the low-frequency part of spectral density functions.

From the linear least square fit of the data, $\alpha = 3.78 \times 10^{-2}$ and $\beta = 2.82 \times 10^{-10}$ (Figure 5), and the corresponding roots at −0.28, 8.65, and 0.77 ns. The first root has no physical meaning, while the second and the third, on the basis of their magnitude, may be assigned to the overall tumbling of the molecule and the internal motion, respectively, at the low-frequency end. If we insert the two roots into the Lorentzian spectral density function of eq 11, we obtain two pairs of $J(0)$ and $J(\omega_N)$ values. These two pairs of values fall on the two ends of the correlation line of Figure 5a (indicated with arrows) and represent the two extreme cases of $J(0)$ that would be consistent with the data. This means that a NH group that would be positioned at the upper right-hand position would have a spectral density function consisting of a pure Lorentzian with a τ, the overall correlation time, of 8.65 ns. A point at the lower left-hand end of the correlation line would indicate a NH group that has a spectral density function consisting of a pure Lorentzian with a τ of 0.77 ns. As all experimental values lie between the two positive roots converted in Lorentzians (Figure 5), they can be expressed with a linear combination of these two Lorentzians. This would be the case if only the values of the spectral density function $J(0)$ and $J(\omega_N)$ would be considered. The negative root does not contribute to this expression.

If we perform the same kind of analysis for the correlation of $\langle J(\omega_H)\rangle$ with $J(0)$, we would ideally expect the same roots if the whole spectral density function would be a linear combination of two Lorentzians only. If we do so for the correlation of $J(\omega_H)$ versus $J(0)$, we obtain three positive roots of 6.95, 1.46, and 0.105 ns. It is reasonable that the first two values, 6.95 and 1.46 ns, again describe the overall and an internal correlation time, while the third root could represent another internal motion. The values of the larger two roots are comparable to those obtained from the correlation of $J(\omega_N)$ with $J(0)$. The fact that the numbers obtained from the two correlation diagrams of Figure 5 do not precisely agree might be caused by exchange contributions that corrupt some of the $J(0)$ values in GAL4(1−65). Chemical exchange effects are more pronounced in GAL4-(1−65) than in other proteins of the size (K. T. Dayie and G. Wagner, unpublished results). We have started to apply this analysis to other proteins and find generally a better agreement between the roots obtained from the two correlation diagrams as shown in Figure 5.

In the above analysis, the spectral density function corresponding to internal motion has been modeled with a Lorentzian. As already stated, because internal motions are fast, we expect this function to be quite flat in the low-frequency range, which is a characteristic of a Lorentzian. However, the calculated internal correlation time should not be considered as a real physical parameter. The overall spectral density function for the internal motion may itself be made of several contributions.

With the proviso given above, the low-frequency part of the spectral density function can thus be expressed at the level of each residue, as a linear combination of only two Lorentzians, each corresponding to one type of motion (overall and internal). These two Lorentzians cross over at 61.6 MHz. This explains the moderate sensitivity of $J(\omega_N)$ to the position of the residue in the protein (Figure 4b), as the nitrogen frequency at a spectrometer field of 11.744 T lies at 50.66 MHz and is very close to this isosbestic point. While we assume a Lorentzian shape at low frequencies, we make no such assumptions at higher frequencies.

Respective Contribution of Overall Tumbling and Internal Motion. From eq 16, where the contributions of the overall tumbling and internal motions are explicitly separated, a system of three equations is obtained for the three frequency values of $J(\omega)$:

$$J(0) = a_0 J_0(0) + \sum_j a_j J_j(0)$$

$$J(\omega_N) = a_0 J_0(\omega_N) + \sum_j a_j J_j(\omega_N) \qquad (21)$$

$$J(\omega_H) = a_0 J_0(\omega_H) + \sum_j a_j J_j(\omega_H)$$

Assuming an isotropic motion of the protein, $J_0(\omega)$ is equivalent to a Lorentzian function (eq 11) with $\tau_c = 8.65$ ns. It can also be assumed that the spectral density function of internal motions is flat at low frequencies, so that

$$\sum_j a_j J_j(0) = \sum_j a_j J_j(\omega_N) \qquad (22)$$

With the above assumptions, three unknowns are left: a_0,

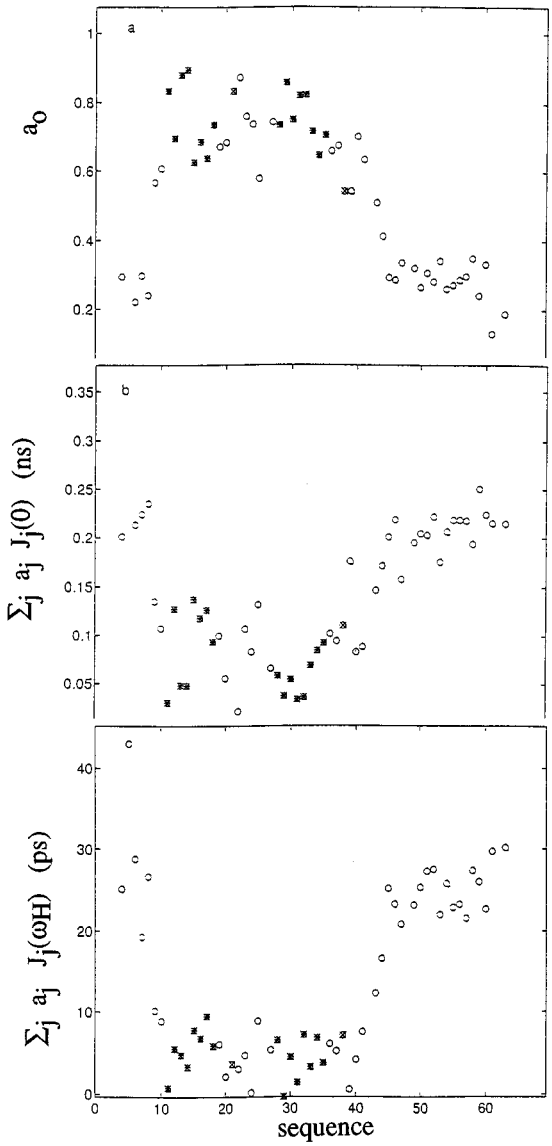

FIGURE 6: Plots versus protein sequence of (a) a_0, the weight of the spectral density function for overall tumbling, (b) $\sum_j a_j J_j(0)$ (in nanoseconds rad), and (c) $\sum_j a_j J_j(\omega_H)$ in picoseconds/rad), the contributions of the spectral density function for internal motions at low and proton frequencies, respectively.

$\sum_j a_j J_j(0)$, and $\sum_j a_j J_j(\omega_H)$. They can be determined for each N−H vector by solving the system of eq 21. The results are shown in Figure 6a−c. As expected, while a_0, the weight of the overall tumbling spectral density function, is small for the N- and the C-termini and large for the core of the protein, the overall contribution of internal motion at low frequencies, $\sum_j a_j J_j(0)$, has the opposite behavior. This is even more pronounced for the weight of internal motions at high frequencies, $\sum_j a_j J_j(\omega_H)$.

Strikingly, a_0 varies significantly within well-defined secondary structures in the core of the protein (Figure 6a). C11, I13, and C14 in the first helix and A29, C31, and L32 in the second one appear to experience a contribution of the overall tumbling motion higher than the average defined by the neighboring residues. This is a consequence of the higher $J(0)$ values attributed to these particular residues, which are most likely affected by fast chemical exchange-type motion (eq 14). The loop between the two helices does not appear flexible except for residue K25 which has a lower a_0 value. Again, two residues (C21 and S22) exhibit a_0 values which are higher than the average, also because of exchange-type motions affecting their $J(0)$ value. In fact, measurements of the relaxation rate in the rotating frame indicate slow conformational exchange processes in the microsecond range at four discontinuous segments of the polypeptide chain, all within the globular portion of the protein. No slow motions on this time scale were found in the long unstructured C-terminal strand or the short undefined N-terminal strand. The four regions exhibiting the slow motions are centered around the four polypeptide regions embracing the metal−thiolate cluster. It is noteworthy that these fluctuations in the slow frequency regime occur around the cysteine residues involved in the zinc cluster. Because the exchange effects are located in and around the cysteine residues which chelate the two zinc atoms, they may be produced by slow isomerization of the cysteine side chains (K. T. Dayie and G. Wagner, unpublished results). Such slow motions have been observed in cysteines engaged in disulfide bonds in BPTI (Szyperski et al., 1993), in residues that are zinc ligands or are adjacent to zinc ligands in the zinc finger DNA binding domain from Xfin (Palmer et al., 1991), and in the first ion binding loop of the Cd^{2+}-loaded calbindin (Akke et al., 1993).

Because the adiabatic contribution of exchange-type motion is not accounted for in the system of eq 21, the values obtained for the spectral density function associated with fast internal motions are slightly corrupted. Where a_0 is larger, $\sum_j a_j J_j(0)$ is smaller than the average. The calculation of $\sum_j a_j J_j(\omega_H)$ is less corrupted. If the internal motions can be described by a single Lorentzian, i.e. if the sums $\sum_j a_j J_j(0)$ and $\sum_j a_j J_j(\omega_H)$ each contain a single term, their values should be identical except for a scaling factor. Inspection of Figure 6 shows that this is approximately true.

The anticorrelation between the contributions of internal and overall motion is seen more clearly in Figure 7a,b, where variations of $\sum_j a_j J_j(0)$ and $\sum_j a_j J_j(\omega_H)$ versus $1 - a_0$ are indicated. Not surprisingly, $\sum_j a_j J_j(0)$ is linearly correlated to $1 - a_0$ with a maximum value obtained from the graph at the point where the quantity $1 - a_0$ equals 1:

$$\max[\sum_j a_j J_j(0)] = J_{int}(0) = 0.294 \text{ ns} \qquad (23)$$

Here, $J_{int}(\omega)$ is the Lorentzian which corresponds to the internal correlation time (0.77 ns) formerly calculated from the third root of eq 20. More interestingly, the linear correlation between $\sum_j a_j J_j(\omega_H)$ and $1 - a_0$ suggests that all residues are affected by motions of very similar rates of internal motions. Correspondingly, the maximum value of $\sum_j a_j J_j(\omega_H)$ deduced from Figure 7b at the point where the quantity $1 - a_0$ equals 1 is 35.7 ps (Figure 7b). This value of 35.7 ps determined from the graph can be compared to one calculated assuming a shape for the spectral density function for the internal motion. First, assume a single Lorentzian for the internal motion with a correlation time of 0.77 ns deduced earlier from eq 20. Second, assume a single Lorentzian for the internal motions so that $\sum_j a_j J_j(\omega_H) = a_1 J_1(\omega_H)$, and without loss of generality, take the limiting case of a_1 at the point where $1 - a_0 = 1$. A simple calculation yields a value of 53 ps. Thus, the two maximum values at low and high frequencies of the internal spectral density function [$\max[\sum_j a_j J_j(0)]$ and $\max[\sum_j a_j J_j(\omega_H)]$] do not exactly fall on the same Lorentzian. This suggests that the spectral density function for the fast internal motions is either

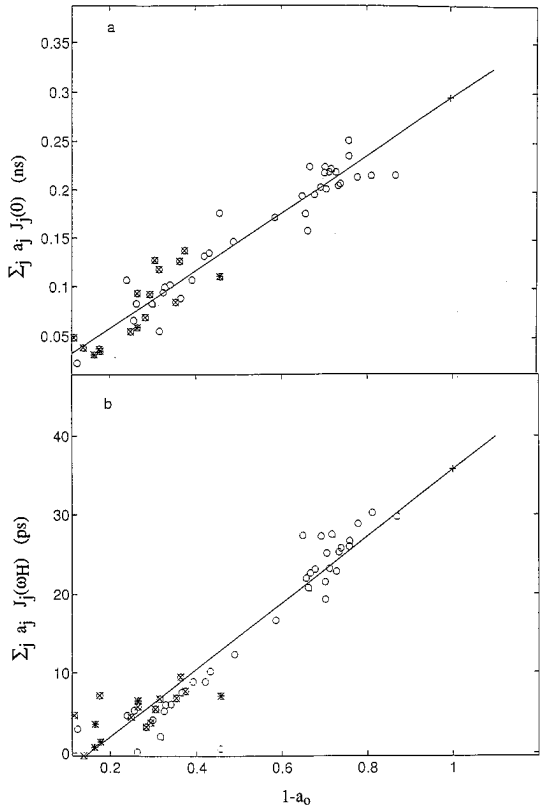

FIGURE 7: Plots of the contribution of internal spectral density function at (a) low and (b) proton frequencies, as a function of $1 - a_0$, the overall weight of spectral density function for internal motions. (+) Points corresponding to the linear fit to $1 - a_0 = 1$.

not a pure Lorentzian or is composed of a sum of Lorentzians corresponding to a wide range of small correlation times, or there is a corruption of the apparent $J(0)$ values due to exchange effects. A full spectral density mapping study at multiple fields without invoking the high-field approximation could help in the detection of such an increase in $J(0)$ due to exchange or nonmonotonic decrease in the values of the spectral density function at the proton frequencies.

Comparison with the Model Free Approach Analysis of the Relaxation Data. The model free approach (Lipari & Szabo, 1982a,b), which has been widely use to analyze the relaxation data in proteins, is based on the following expression for the spectral density function:

$$J(\omega)_{1s} = {}^2/_5\left(\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S^2)\tau_e}{1+(\omega\tau_e)^2}\right) \qquad (24)$$

Here, the order parameter S^2 is related to the amplitude of the rapid internal movements, τ_m is the overall rotational correlation time for molecular tumbling, and τ_e is the harmonics of the overall and the internal correlation time which pertains to each residue: $\tau_e^{-1} = \tau_m^{-1} + \tau_i^{-1}$. A physical interpretation of the τ_i values requires a more specific model of motion, such as the wobbling in a cone (Woessner, 1962; Kinoshita et al., 1977; Richarz et al., 1980) or the jump approach (Tropp, 1980).

The coefficients a_0 of the present study are analogous to the order parameter S^2. The three unknowns, S^2, τ_m, and τ_e, can be determined from the three experimental values of $J(\omega)$ by a simple least square fit for each residue. When the calculation is done independently for each residue, τ_m is

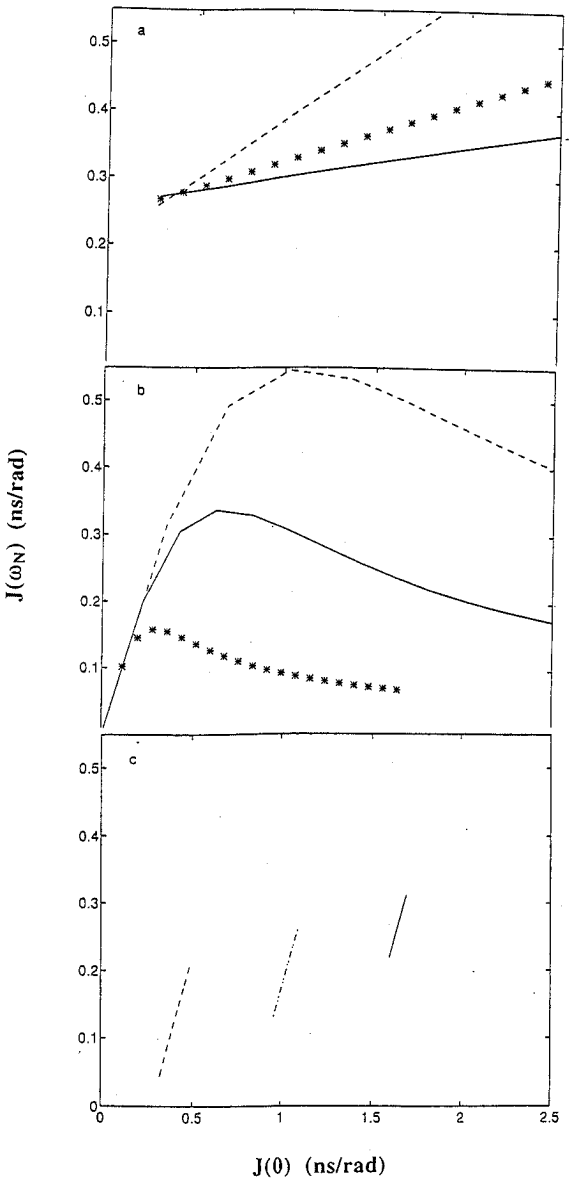

FIGURE 8: Simulation of possible correlations of $J(\omega_N)$ versus $J(0)$ as functions of parameters from eq 24. (a) Variation of S^2 from 0 to 1, $\tau_e = 0.777$ ns, $\tau_m = 5$ ns (- - -), 7 ns (***), and 8.65 ns (—). (b) Variation of τ_m from 0 to 20 ns, $\tau_e = 0.1$ ns, $S^2 = 0.85$ (—), 0.5 (***), and 0.2 (- - -). (c) Variation of τ_e from 0 to 0.5 ns, $\tau_m = 8$ ns, $S^2 = 0.9$ (—), 0.5 (***), and 0.3 (- - -). In all simulations, a spectrometer field of 11.744 T is assumed (500 MHz proton frequency).

found to vary along the sequence, around an average value of 5.5 ns in the terminal ends and between 7 and 9 ns in the core, and the order parameters, S^2, are significantly larger than the coefficients a_0 (data not shown). When an overall correlation time of 8.65 ns is imposed on all residues, S^2 and a_0 behave very similarly. The τ_i values however do not relate with the internal spectral density functions $\sum_j a_j J_j(0)$ or $\sum_j a_j J_j(\omega_H)$ (data not shown).

The Lipari−Szabo method is inappropriate in cases where there is no well-defined overall correlation time as may occur in molecules with little persistent structure in solution. A limited number of such studies have appeared in the literature (van Mierlo et al., 1993; Alexandrescu & Shortle, 1994). Unlike the use of the spectral densities described in the previous section, the order parameters are sensitive to only nanosecond or faster motions. Motions comparable to or

slower than the overall tumbling are not accessible by this method. In such instances, the discussion of the dynamics in terms of the raw rates might be more profitable or the use of spectral density functions as reported here and by Farrow et al. (1995) more appropriate.

To gain some independent insight into the correlation between $J(\omega_N)$ and $J(\omega_H)$ with $J(0)$, we simulated the raw data based on eq 24. We wanted to see whether the linear relation between $J(\omega_N)$ and $J(0)$ can be related to predominantly a single parameter of eq 24. We carried out the simulations under the following three conditions. (i) We vary τ_e and keep the order parameter and τ_m constant. This simulation is shown in Figure 8c. Obviously, this does not yield the correlation as seen in Figure 5a. (ii) We vary the correlation time for the overall tumbling and keep the order parameter S^2 and τ_e constant. This is shown in Figure 8b; it also does not show any relation to the correlation shown in Figure 5a. Finally, (iii) we vary the order parameter and fix the overall correlation time and τ_e, and the results are shown in Figure 8a. The resulting plots of $J(\omega_N)$ and $J(\omega_H)$ versus $J(0)$ reproduce the linear correlation seen experimentally (Figure 5a). This result suggests that the data points are ordered by the degree of libration (the variation of the order parameter) along the protein sequence. The forces acting on the backbone nuclei are collective; residues not involved in such collective motion show large deviation from linearity. Thus, the raw spectral density functions contain all the information about the detailed types of motion along the backbone. The slope of this correlation map is determined by the overall tumbling; in other words, all nuclei along the backbone are subjected to an overwhelming overall diffusive thermal agitation with the protein matrix and the solvent.

CONCLUSIONS

We have demonstrated the feasibility and utility of reduced spectral density mapping using a subset of the six relation rates to analyze the frequency spectrum of the internal dynamics of a protein. The three traditionally measured rates [$R_N(N_z)$, $R_N(N_{x,y})$, and $R_N(H_z \rightarrow N_z)$] allow the calculation of the spectral density functions at 0, ^{15}N, and ^1H frequencies without the complication of errors propagated at the high frequencies. Addition of a fourth relaxation term, the two-spin J-ordered rates ($2N_zH_z$), leads to the same results together with the longitudinal relaxation rate of the proton, ρ_H, which is difficult to obtain by a direct measurement. This determination of ρ_H is more precise than in the original spectral density mapping procedure (Peng & Wagner, 1992a,b) because now this value is obtained from a linear combination of only three rates (see eq 5a) which all can be measured rather reliably, compared to a combination of five rates as done previously (Peng & Wagner, 1992a,b).

The spectral density functions delineate the spectrum of frequencies available to the secondary structural elements of the protein. The power of the motion given by the intensity of the spectral density function is gathered toward low frequencies in the core of the protein, while it is more spread toward high frequencies in the unstructured termini parts. This is due to the fact that the overall motion of the protein has a larger contribution in the central part of the protein than in the ends. The analysis deduced from the linear correlation between $J(\omega_N)$ and $J(0)$ points to the fact that the spectral density function attached to the internal motions, as measured at 0 and ω_N, appears very similar from one residue to the other. In other words, variations of $J(0)$ and $J(\omega_N)$ observed along the sequence are mainly due to different weightings of two spectral density functions which appears to be common for all residues: one corresponding to the overall tumbling and the other to a dominant (relatively slow) type of internal movements. The latter cannot be reduced to a single Lorentzian, suggesting that it is itself composed of various contributions. This is corroborated by the fact that the correlation of $J(\omega_H)$ with $J(0)$ seems to indicate a contribution of more than two types of motions. The contributions to the spectral density function at ω_H of the overall motion and the dominant type of internal motions are so low that faster internal motions become significant at the high frequency. This simple decomposition of the spectral density function has been observed with other proteins (data not shown), and relaxation data are currently being reinvestigated in order to analyze the dependence of the internal spectral density function on size and type of the protein as well as on temperature.

Finally, this analysis also reveals the presence of slow motions which are reflected in the high values of the spectral density function at zero frequency. These slow motions affect residues in the vicinity of the cysteines engaged in the zinc cluster. Measurements of relaxation in the rotating frame have indicated motions in the millisecond range (K. T. Dayie and G. Wagner, unpublished results). Thus, the apparent value of $J(0)$ contains a contribution from internal motions covering a wide frequency range from picoseconds to milliseconds.

ACKNOWLEDGMENT

We thank Professors Steve Harrison and Mark Ptashne for the clone that expresses GAL4(1−65) and Drs. Marmorstein and Jim Baleja and Ted Mau for the sample of GAL4-(1−65) used in the study.

SUPPORTING INFORMATION AVAILABLE

Tables showing (1) the results of a statistical analysis of the reduced spectral density function approach using three, four, five, and six rate parameters; (2) $R_N(N_z)$, $R_N(N_x)$, NOE, $R_{NH}(2H_zN_z)$, $R_{NH}(2H_zN_{x,y})$, and $R_H(H_z)$ experimental relaxation rate constants at 500 MHz; and (3) the reduced spectral density values at 0, 50, and 450 MHz (7 pages). Ordering information is given on any current masthead page.

REFERENCES

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

Akke, M., Skelton, N. J., Kördel, J., Palmer, A. G., III, & Chazin, W. J. (1993) *Biochemistry 32*, 9832−9844.

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

Baleja, J. D., Marmorstein, R., Harrison, S. C., & Wagner, G. (1992) *Nature 356*, 450−453.

Bloom, M., Reeves, L. W., & Wells, E. J. (1965) *J. Chem. Phys. 42*, 1615−1624.

Bram, R., & Kornberg, R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 43−47.

Brooks, C. L., III, Karplus, M., & Pettitt, B. M. (1987) *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, pp 1−259, John Wiley & Sons, New York.

Carey, M., Kakidani, H., Leatherwood, J., Mostashari, F., & Ptashne, M. (1989) *J. Mol. Biol. 209*, 423−432.

Deverell, C., Morgan, R. E., & Strange, J. H. (1970) *Mol. Phys. 18*, 553−559.
Farrow, N. A., Zhang, O., Forman-Kay, J. D., & Kay, L. E. (1995a) *Biochemistry 34*, 868−878.
Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., & Kay, L. E. (1995b) *J. Biomol. NMR 6*, 153−162.
Fedor, M. J., Lue, N. F., & Kornberg, R. D. (1988) *J. Mol. Biol. 204*, 109−127.
Frauenfelder, H., Parak, F., & Young, R. D. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 451−479.
Giniger, E., Varnum, S. M., & Ptashne, M. (1985) *Cell 40*, 767−774.
Hiyama, Y., Niu, C., Silverton, J. V., Bavoso, A., & Torchia, D. A. (1988) *J. Am. Chem. Soc. 110*, 2378−2383.
Ishima, R., & Nagayama, K. (1995a) *Biochemistry 34*, 3162−3171.
Ishima, R., & Nagayama, K. (1995b) *J. Magn. Reson., Ser. B 108*, 73−76.
Johnston, M. (1987a) *Microbiol. Rev. 51*, 458−476.
Johnston, M. (1987b) *Nature 328*, 353−355.
Karplus, M., & McCammon, J. A. (1983) *Annu. Rev. Biochem. 53*, 263−300.
Kegan, L., Gill, G., & Ptashne, M. (1986) *Science 231*, 699−704.
Keiter, E. A. (1986) Ph.D. Thesis, University of Illinois.
Kinoshita, K., Kawato, S., Jr., & Ikegami, A. (1977) *Biophys. J. 20*, 289−305.
Kraulis, P. J., Raine, A. R. C., Gadhavi, P. L., & Laue, E. D. (1992) *Nature 356*, 448−450.
Lin, Y.-S., Carey, M., & Ptashne, M. (1988) *Cell 54*, 659−664.
Lipari, G., & Szabo, A. (1982a) *J. Am. Chem. Soc. 104*, 4546−4559.
Lipari, G., & Szabo, A. (1982b) *J. Am. Chem. Soc. 104*, 4559−4570.
London, R. E., & Avitable, J. (1978) *J. Am. Chem. Soc. 100*, 7159−7165.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Marmorstein, R., Carey, M., Ptaschne, M., & Harrison, S. C. (1992) *Nature 356*, 408−414.
Palmer, A. G., Rance, M., & Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371−4380.
Peng, J. W., & Wagner, G. (1992a) *J. Magn. Reson. 98*, 308−332.
Peng, J. W., & Wagner, G. (1992b) *Biochemistry 31*, 8571−8586.
Peng, J. W., & Wagner, G. (1995) *Biochemistry 34*, 16733−16752.
Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vertterlinh, W. T. (1988) *Numerical Recipes in C- The Art of Scientific Computing*, Cambridge University Press, New York.
Richarz, R., Nagayama, K., & Wüthrich, K. (1980) *Biochemistry 19*, 5189−5196.
Sandström, J. (1982) *Dynamic NMR Spectroscopy*, Academic Press, New York.
Shirakawa, M., Fairbrother, W. J., Serikawa, Y., Ohkubo, T., Kyogoku, Y., & Wright, P. E. (1993) *Biochemistry 32*, 2144−2153.
Szyperski, T., Luginbühl, P., Otting, G., Güntert, P., & Wüthrich, K. (1993) *J. Biomol. NMR 3*, 151−164.
Tropp, J. (1980) *J. Chem. Phys. 47*, 6035−6043.
Tycko, R., Ed. (1994) *Nuclear Magnetic Resonance Probes of Molecular Dynamics*, Kluwer Academic Publishers, The Netherlands.
van Mierlo, C. P. M., Darby, N. J., Keeler, J., Neuhaus, D., & Creighton, T. E. (1993) *J. Mol. Biol. 229*, 1125−1146.
Wagner, G. (1993) *Curr. Opin. Struct. Biol. 3*, 748−754.
Wagner, G. (1994) Approaches for Studies of Protein Mobility and Structure, The 35th Experimental NMR Conference, April 10−15, 1994, Asilomar, CA.
Wittebort, R. J., & Szabo, A. (1978) *J. Chem. Phys. 69*, 1723−1736.
Woessner, D. E. (1962) *J. Chem. Phys. 36*, 647−654.

BI9526802