Structural and Dynamical Characterization of a Biologically Active Unfolded Fibronectin-Binding Protein from *Staphylococcus aureus*†

Christopher J. Penkett,‡ Christina Redfield,‡ Jonathan A. Jones,‡ Ian Dodd,§ Julia Hubbard,|| Richard A. G. Smith,§,‡ Lorna J. Smith,‡ and Christopher M. Dobson*,‡

Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, OX1 3QT, U.K., AdProTech plc, Unit 3, 2 Orchard Road, Royston SG8 5HD, U.K., and SmithKline Beecham Pharmaceuticals, New Frontiers Science Park, North Site, Third Avenue, Harlow, Essex CM19 5AW, U.K.

Received June 15, 1998; Revised Manuscript Received September 16, 1998

ABSTRACT: A 130-residue fragment (D1−D4) taken from a fibronectin-binding protein of *Staphylococcus aureus*, which contains four fibronectin-binding repeats and is unfolded but biologically active at neutral pH, has been studied extensively by NMR spectroscopy. Using heteronuclear multidimensional techniques, the conformational properties of D1−D4 have been defined at both a global and a local level. Diffusion studies give an average effective radius of 26.2 ± 0.1 Å, approximately 75% larger than that expected for a globular protein of this size. Analysis of chemical shift, $^3J_{HN\alpha}$ coupling constant, and NOE data show that the experimental parameters agree well overall with values measured in short model peptides and with predictions from a statistical model for a random coil. Sequences where specific features give deviations from these predictions for a random coil have however been identified. These arise from clustering of hydrophobic side chains and electrostatic interactions between charged groups. ^{15}N relaxation studies demonstrate that local fluctuations of the chain are the dominant motional process that gives rise to relaxation of the ^{15}N nuclei, with a persistence length of approximately 7−10 residues for the segmental motion. The consequences of the structural and dynamical properties of this unfolded protein for its biological role of binding to fibronectin have been considered. It is found that the regions of the sequence involved in binding have a high propensity for populating extended conformations, a feature that would allow a number of both charged and hydrophobic groups to be presented to fibronectin for highly specific binding.

The extracellular matrix (ECM)[1] consists of a complex mixture of glycoproteins and proteoglycans (*1*) that both provides a structural framework for the tissue and affects the cellular physiology of the organism. Many pathogenic microorganisms use cell surface adhesins to target ECM macromolecules as an initial step for invading host cells (*2*). The acronym MSCRAMM (microbial surface components recognizing adhesive matrix molecule) has been used to describe this family of microbial adhesins. Fibronectin, a dimeric mosaic protein composed of modular protein units that often correspond to the exon structure of the gene (*3*), was the first ECM protein shown to act as a substrate for microbial adhesion (*4*). Three genes corresponding to fibronectin-binding proteins in *Staphylococcus aureus* have been identified (encoding for the proteins FnBPA, FnBPB, and FnBPC; *5*−*7*), and fibronectin-binding proteins have also been found in group A, C, and G streptococci (*8*).

The sequences of these microbial fibronectin-binding proteins have common features. As shown for *S. aureus* fibronectin-binding protein FnBPA in Figure 1a, at the N-terminus there is a signal peptide (S) and a large segment of sequence unique to the receptor (U). Following this is the primary fibronectin-binding site consisting of a motif of about 30−40 amino acids, repeated two to five times; in the three *S. aureus* fibronectin-binding proteins, this ligand-binding site has been located in a series of three repeats of 38−39 residues (D1, D2, and D3; see Figure 1b) and a partial fourth repeat (15 residues; D4) that can each bind fibronectin with low affinity and that in tandem (D1−D4) comprise a high-affinity fibronectin-binding domain (*5*). C-terminal to this fibronectin-binding domain are features characteristic of cell-surface-associated proteins from Gram-positive bac-

† This is a contribution from the Oxford Centre for Molecular Sciences, which is supported by the U.K. Engineering and Physical Sciences Research Council, the Biotechnology and Biological Sciences Research Council, and the Medical Research Council. L.J.S. is a Royal Society University Research Fellow. The research of C.M.D. is supported in part by an International Research Scholars award from the Howard Hughes Medical Foundation and by the Wellcome Trust.

* To whom correspondence should be addressed.

‡ University of Oxford.

§ AdProTech plc.

|| SmithKline Beecham Pharmaceuticals.

[1] Abbreviations: 2D and 3D, two- and three-dimensional; CD, circular dichroism; ECM, extracellular matrix; HMQC, heteronuclear multiple-quantum correlation; HSQC, heteronuclear single-quantum correlation; MSCRAMM, microbial surface components recognizing adhesive matrix molecule; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement and exchange spectroscopy; sc, side chain; ppm, parts per million; $R_1 = (T_1)^{-1}$, $R_2 = (T_2)^{-1}$, $R_{1\rho} = (T_{1\rho})^{-1}$, longitudinal, transverse, and rotating frame relaxation rates; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; TOCSY, total correlation spectroscopy; $^3J_{HN\alpha}$, scalar coupling constant between H^N and H^α; αN-(i,i+1), NOE cross-peak between H^α of residue i and H^N of residue i+1.

10.1021/bi9814080 CCC: $15.00
Published on Web 11/11/1998

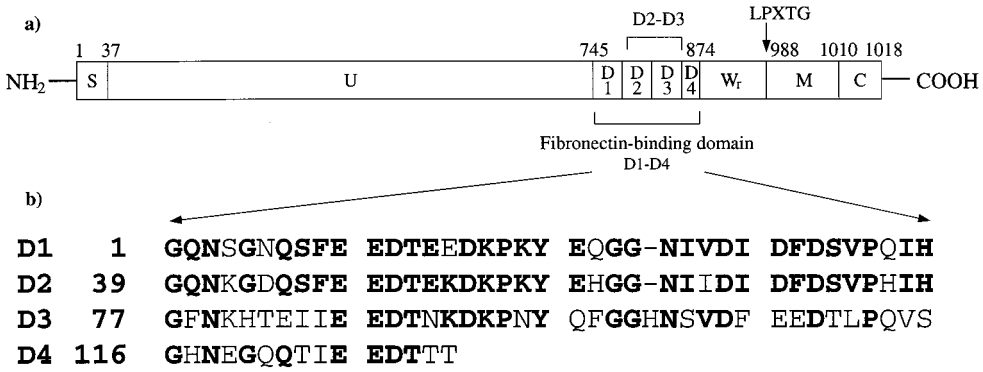

FIGURE 1: (a) Domain organization of the fully sequenced *S. aureus* fibronectin-binding protein FnBPA (*5*) showing the four fibronectin-binding repeats D1, D2, D3, and D4. S, signal sequence; U, sequence unique to the fibronectin-binding protein; W_r, cell-wall-spanning region; M, membrane-spanning region; C, intracellular sequence. The residue numbers for the domain boundaries are indicated above the schematic diagram. (b) Sequence alignment of D1, D2, D3, and D4 from *S. aureus* fibronectin-binding protein FnBPC (*7*) highlighting conserved residues in bold. The residue numbers for the D repeats are also indicated. D4 contains an additional residue at the C-terminus, Thr-130, which is a cloning artifact.

teria. These include a cell-wall-spanning segment rich in proline residues (W_r); the sequence LPXTG, involved in sorting the protein to the cell wall (*9*); a hydrophobic membrane-spanning segment (M); and a short cytoplasmic segment (C) consisting of positively charged amino acids.

Studies by CD spectroscopy and other biophysical techniques have indicated that the fibronectin-binding domains from various microbial fibronectin-binding proteins contain very little secondary structure in phosphate buffer at neutral pH (*10*, *11*). We have also recently shown using NMR methods to study the local conformational properties that the fibronectin-binding domain from *S. aureus*, D1−D4, is largely unstructured in aqueous solution at pH 6.0 (*12*). Despite its unfolded nature, however, D1−D4 is biologically active. It has significant affinity for fibronectin and its fragments when these proteins are bound to solid surfaces, and it also antagonizes the binding of *S. aureus* strains expressing FnBPA to surfaces coated with fibronectin. The latter property suggests a possible utility for this class of polypeptide in preventing colonization of implanted medical devices by bacteria whose adherence to foreign surfaces is fibronectin-mediated (*7)* (I. A. Critchley, K. May, G. Gardiner, D. McBay, I. Dodd, and G. Smith, unpublished results).

Here we report a detailed study of D1−D4 using heteronuclear NMR techniques coupled with predictions from a statistical model of a random coil. Extraction of chemical shifts, coupling constants, and NOEs has enabled the local conformational properties of the protein to be defined while diffusion studies have been used to probe the global characteristics through measurement of the effective hydrodynamic radius. In addition, ^{15}N relaxation rates have been used to probe the dynamics of the polypeptide chain. Through this work, we have obtained a detailed description of the conformational ensemble defining D1−D4, an unfolded biologically active protein, in the absence of denaturants. One interesting aspect of the work is that the structural and dynamical properties of D1−D4 (130 residues) are compared with those of a smaller fragment D2−D3 (77 residues) that consists of the two repeats D2 and D3 from D1−D4. One other feature of interest with the present study relates to the high-sequence homology for the different D repeats of D1−D4 [there is between 44% and 84% sequence identity between D1, D2, and D3 (see Figure 1b), while D4 is shorter in length but has 53% identity with the first 15 residues of D1], which allows comparison of the NMR parameters between the different D repeats.

MATERIALS AND METHODS

Sample Preparation. A gene encoding the D-repeat region of fibronectin-binding protein (FnBP) from a clinical isolate of *S. aureus* (J2385 strain B) was cloned by PCR and sequenced (*7*). This fragment was found to be almost identical to the homologous regions within the *S. aureus* 8325-4 FnBPA and FnBPB genes (*5*, *6*), although some coding differences were observed, notably N752D, S803N, K821Q, Q825H, and a four amino acid deletion (P838−V841) as compared with FnBPA. In view of these differences, the *S. aureus* J2385 protein has been denoted type C. Both D1−D4 and the shorter derivative D2−D3 were expressed from *Escherichia coli* BL21 (DE3) using the vector pBROC413 (*13*), itself a derivative of pT7-7 (*14*). Both proteins were purified to greater than 95% homogeneity from the soluble lysate fraction using affinity chromatography on Sepharose-bound fibronectin [the latter having been purified from a human plasma fraction by the method of Miekka et al. (*15*)]. Alternatively, ion exchange on Q-Sepharose followed by high-resolution hydrophobic interaction chromatography on toyopearl butyl was employed. D1−D4 had a molecular mass of 19 970 (expected 19 969) by electrospray mass spectrometry although the apparent molecular mass under reducing and nonreducing SDS−PAGE conditions was significantly higher. Fully ^{15}N-labeled samples of D1−D4 and D2−D3 were obtained by means of similar procedures using a fermentation medium with $^{15}NH_4Cl$ as the sole nitrogen source and were found to have greater than 97% incorporation of ^{15}N by electrospray mass spectrometry. Both D1−D4 and D2−D3 are inherently stable molecules and retained anti-adhesion activity after exposure to acidic solvents, which enables them to be formulated by lyophilization from 50 mM formic acid. This method was used for the preparation of all samples for NMR studies.

NMR Spectroscopy. NMR samples were prepared to contain approximately 1−2 mM protein by dissolving the proteins in 90%/10% H_2O/D_2O. The pH was adjusted to pH 6.0 with sodium hydroxide and hydrochloric acid. All

the NMR experiments were performed at 5 °C (unless stated otherwise) on home-built spectrometers belonging to the Oxford Centre for Molecular Sciences with ^{1}H operating frequencies of 500.1, 600.2, and 750.1 MHz.

2D ^{15}N−^{1}H HSQC (*16*), 3D NOESY−HSQC (*17-19*), 3D TOCSY−HSQC (*18*, *20*), and 3D HSQC−NOESY−HSQC (*21*, *22*) experiments using gradient sensitivity-enhanced methods (*23*) were recorded with mixing times of 200 and ∼50 ms in the NOESY and TOCSY experiments, respectively. For the NOESY−HSQC and TOCSY−HSQC experiments, the data sets typically consisted of 128, 32, and 512 complex points in t_1, t_2 and t_3, respectively. The data sets for the HSQC−NOESY−HSQC experiments comprised 64, 40, and 512 complex points in t_1, t_2 and t_3, respectively. All 3D spectra were resolution enhanced in t_3 by a double-exponential multiplication (DM) and in t_1 and t_2 by shifted (72°) squared sine-bell windows.

Diffusion coefficients for D1−D4 and D2−D3 were determined by pulsed field gradient spin−echo NMR (*24*) using PG-SLED pulse sequences (*25*). A small quantity of dioxan was added to each sample as a radius standard (*26*) allowing values of effective radii to be determined.

$^3J_{HN\alpha}$ coupling constants were measured from 2D ^{15}N−^{1}H HMQC−*J* experiments (*27*, *28*) recorded at a ^{1}H frequency of 750 MHz. The data sets comprised 512 complex t_1 increments of 1024 complex data points. In t_2, a DM window was used to enhance the resolution, while in t_1 a shifted (30°) sine-bell window was applied. The data sets were zero-filled to 2048 real data points in t_1, giving a final digital resolution of 0.65 Hz/point in F_1. Cross sections for non-glycine residues were simulated by applying the above window function to in-phase Lorentzian doublets in which both components have different line widths. These were then fitted to experimental cross sections parallel to F_1 through cross-peaks of the HMQC−*J* spectra by optimization of the coupling constant and line widths using in-house software (*29*).

Longitudinal ($R_1 = 1/T_1$), transverse ($R_2 = 1/T_2$), and rotating frame ($R_{1\rho} = 1/T_{1\rho}$) relaxation rates and heteronuclear NOE effects ({^{1}H}−^{15}N NOEs) of ^{15}N nuclei of D1−D4 and D2−D3 were measured at a ^{1}H frequency of 600 MHz using procedures described previously (*30*−*32*). Data sets were acquired using between 96 and 128 complex t_1 increments of 1024 complex data points. The recycle delay was set to 2.0 s in all experiments except for the heteronuclear NOE experiment in which the recycle delay was set to 4.0 s. For the R_1 measurements, a series of 9 or 10 spectra with relaxation delays ranging from 20 to 2005 ms was carried out. For the various different $R_{1\rho}$ and R_2 measurements, 10 different relaxation delays were used ranging from 8.9 to 445 ms. A CPMG (Carr−Purcell−Meiboom−Gill) duty cycle delay of 0.5 ms was used in the R_2 experiments. For the $R_{1\rho}$ experiments, a power of 0.7 kHz was used for the ^{15}N spin lock and a power of 2.2 kHz for proton decoupling; coherence transfer due to scalar coupling between ^{1}H and ^{15}N spins can be neglected at these spin-locking field strengths (*33*). $R_{1\rho}$ relaxation rates were corrected for offset effects using the measured R_1 relaxation rates.

The R_1, R_2, $R_{1\rho}$, and NOE data sets were processed using the DM filtering function in t_2 and a shifted (90°) sine-bell function in t_1 and were zero-filled to 2048 and 512 real data points in t_2 and t_1, respectively. Peak heights rather than integrated peak volumes were used to determine relaxation rates and heteronuclear NOE effects, since accurate integration requires very well resolved resonances (*34*). Relaxation rates were fitted as single-exponential decays using software from MODELFREE 3.1 (A. G. Palmer, unpublished results). The procedures used for obtaining errors in the relaxation rates were similar to those employed by Palmer et al. (*35*). The heteronuclear NOE effects were calculated as the ratios of peak heights in the spectrum recorded with proton saturation to those in the spectrum recorded without saturation. The standard deviations of the heteronuclear NOE value were determined on the basis of measured background noise levels using Monte Carlo simulations (*36*).

Calculation of Random Coil NMR Parameters. The prediction of $^3J_{HN\alpha}$ coupling constants and NOE intensities for a random coil used torsion-angle distributions taken from a database of 85 high-resolution crystal structures in the Protein Data Bank (PDB) (*37*, *38*). $^3J_{HN\alpha}$ coupling constants were calculated from these distributions as population weighted averages using the ϕ distribution for each amino acid and the Karplus equation with the parameters of Pardi et al. (*39*). This assumes that there is rapid interconversion between conformers in the random coil ensemble so NMR parameters will be averaged. NOE intensities were calculated using a simplified model of the polypeptide chain and a Monte Carlo procedure to generate ensembles of conformers from the residue-specific torsion angle distributions as described previously (*33*, *40*). For comparison with the experimental data, only NOEs with predicted intensities greater than that expected for a fixed inter-proton distance of 4.4 Å were included; this distance was chosen on the basis of the experimental signal-to-noise levels as described previously (*33*). For NOEs involving degenerate resonances (either glycine H$^\alpha$ groups or side-chain methyl groups), this distance was increased by 0.4 to 4.8 Å (*33*).

Secondary Structure Predictions. Secondary structure predictions (*41*, *42*) for D1−D4 were performed using software from the Genetics Computer Group (GCG) and neural network algorithms (*43*, *44*) from the European Molecular Biology Laboratory (EMBL).

RESULTS AND DISCUSSION

Resonance Assignments, $^3J_{HN\alpha}$ Coupling Constants, and NOEs

Spectral Assignment. Assignment of the ^{1}H and ^{15}N resonances of both D1−D4 and D2−D3 at pH 6.0 and 5 °C was carried out using a combination of 3D NOESY−HSQC, 3D TOCSY−HSQC, and 3D HSQC−NOESY−HSQC experiments. In the spectra of D1−D4 there is a large amount of resonance overlap, the limited ^{1}H chemical shift dispersion reflecting the unfolded nature of the protein; this severely hampered the sequential assignment process. In addition, difficulties arose from the high sequence homology between the different D repeats, the sequence of D2, for example, differing from that of D1 in only six places. First, the repeating sequence means that some spin-system connectivities occur many times; for example, the sequence EEDT occurs five times in the sequence of D1−D4 (Glu-10 to Thr-13, Glu-48 to Thr-51, Glu-86 to Thr-89, Glu-107 to Thr-

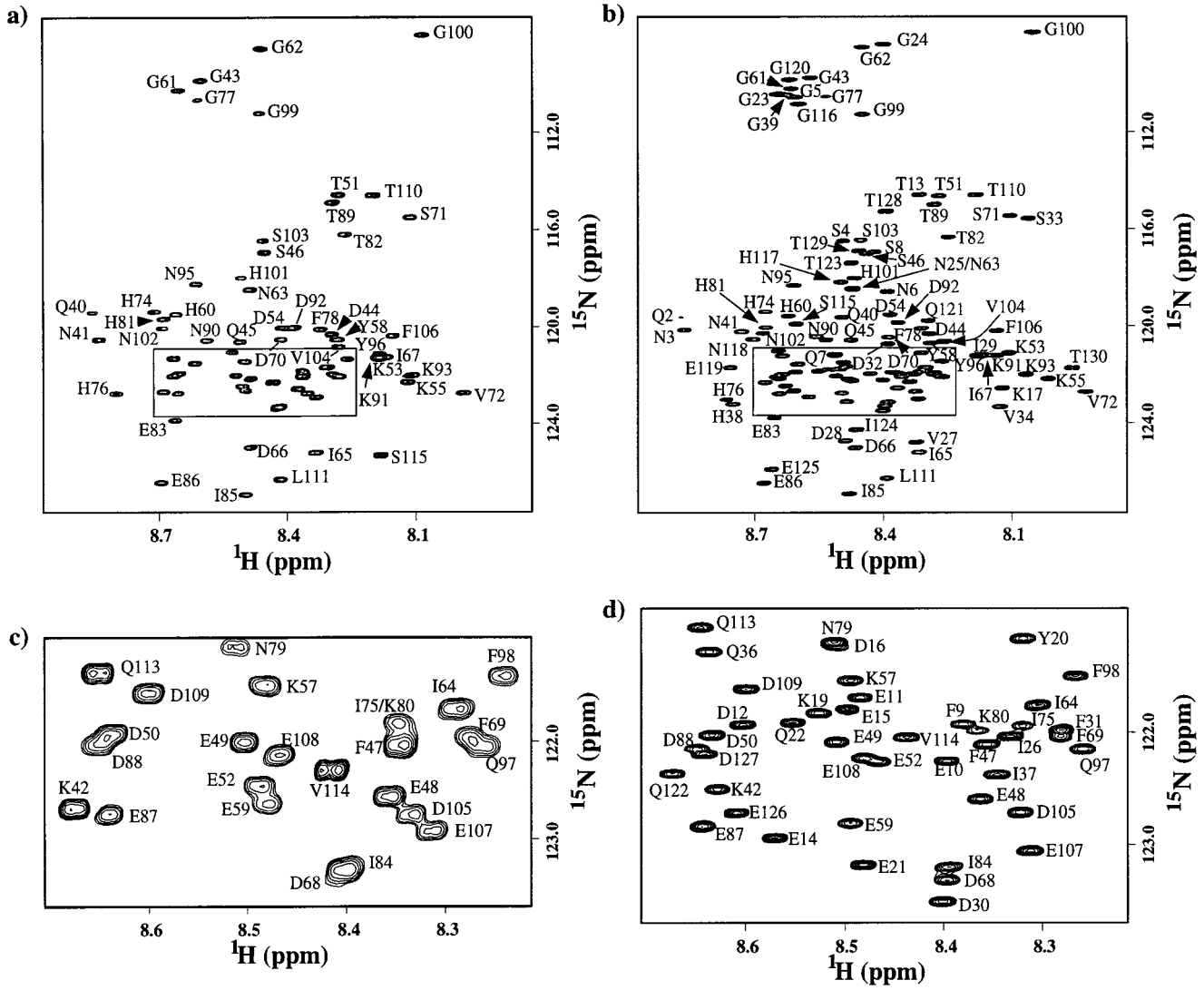

FIGURE 2: (a and b) Fingerprint region and (c and d) expansion of 2D ^{15}N−^{1}H HSQC spectra of (a and c) D2−D3 and (b and d) D1−D4 at pH 6.0 and 5 °C; the backbone cross-peaks are labeled. These spectra were recorded at 500 MHz, except for the spectrum in panel d, which was recorded at 750 MHz to overcome the increased overlap present in this region of the HSQC spectrum of D1−D4.

110, and Glu-125 to Thr-128). Second, the repeating sequence gives rise to an effective reduction in the dispersion of the ^{15}N chemical shift in some regions of the ^{15}N−^{1}H HSQC spectrum of D1−D4. This is because most of the ^{15}N dispersion in unfolded states arises from the identity of the preceding [or (*i*−1)] residue [the ^{15}N chemical shift of residue *i* changes by up to 4.7 ppm depending on the type of residue at position (*i*−1) (*45*)]. The strategy devised to overcome these sequence-related problems was to assign initially the spectra of D2−D3 and then transfer these assignments to the spectra of D1−D4.

Spin systems for the fingerprint peaks of the HSQC spectra of D2−D3 and D1−D4 were obtained from 3D TOCSY−HSQC spectra. Sequential assignments for cross-peaks in the D2−D3 HSQC spectrum were then obtained using sequential αN(*i*,*i*+1) and NN(*i*,*i*+1) NOEs in 3D NOESY−HSQC and HSQC−NOESY−HSQC spectra. A significant number of the cross-peaks in the D1−D4 HSQC spectrum could then be assigned sequence specifically using the assignments made for the cross-peaks in the D2−D3 HSQC spectrum. This was done first for cross-peaks in the D1−D4 HSQC spectrum that were close to the positions of cross-peaks in the corresponding HSQC spectrum of D2−D3 (using spectral overlay methods; see Figure 2) and second for peaks that were identified as being from the same spin systems as the corresponding cross-peaks of the D2−D3 HSQC spectrum. The 3D NOESY−HSQC and HSQC−NOESY−HSQC spectra for D1−D4 were then used to confirm these assignments using sequential αN(*i*,*i*+1) and NN(*i*,*i*+1) NOEs. The sequence-specific method was then applied to the unassigned peaks, which were primarily in the two repeats D1 and D4.

The assigned fingerprint regions of 2D HSQC spectra of D1−D4 and D2−D3 are shown in Figure 2; tables of ^{1}H and ^{15}N backbone and side-chain assignments for the proteins are given in Supporting Information. Except for cases of severe overlap, ^{1}H chemical shifts were extracted from 2D homonuclear spectra since the indirect dimensions of 3D spectra do not contain sufficient resolution to obtain accurate chemical shift measurements (*46*). ^{15}N chemical shifts for D1−D4 and D2−D3 were measured in high-resolution 2D HSQC spectra.

NMR Parameters. $^{3}J_{HN\alpha}$ coupling constants were measured for D1−D4 and D2−D3 as described previously (*12*).

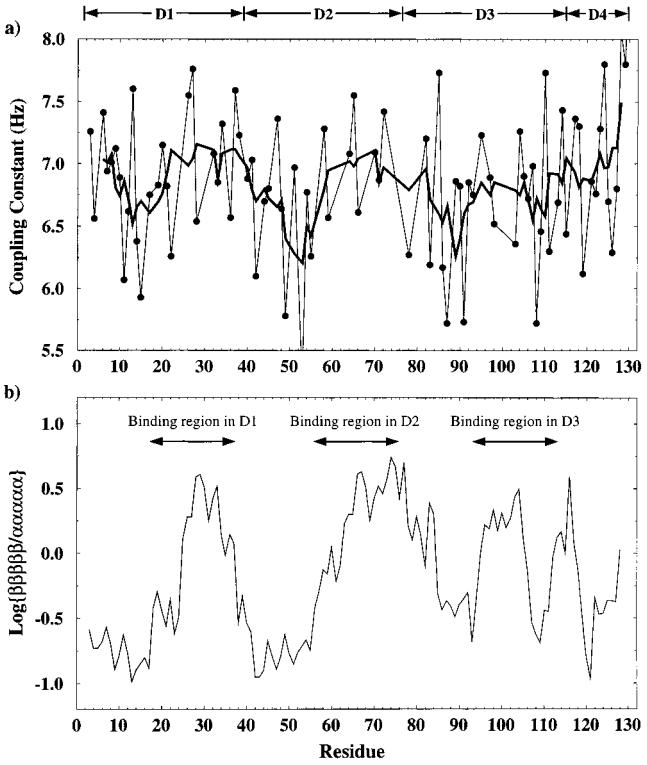

FIGURE 3: (a) Experimental $^3J_{HN\alpha}$ coupling constants for D1−D4. The coupling constants are shown both as individual points and averaged over a sliding window of five residues. (b) The probabilities of all the residues in five-residue segments of the D1−D4 sequence populating α or β main-chain ϕ,ψ conformations. The prediction is from the statistical model for a random coil that takes into account the identity of the preceding residue (*12*). The logarithm ($\log_{10}$) of the all β to all α probability ratio is shown plotted at position (i+2) for each five-residue segment i to (i+4). The fibronectin-binding regions are also indicated.

In Figure 3a, experimental $^3J_{HN\alpha}$ coupling constants for D1−D4 are plotted both as individual points and averaged over a sliding window of five residues for the sequence of D1−D4. A summary of all short- and medium-range NOEs (shown as closed bars) observed in NOESY spectra of D1−D4 is given in Figure 4; in many cases methyl groups are better resolved than β-protons, and hence NOEs from all side-chain (sc) protons are included together (*33*). A similar pattern of short- and medium-range NOEs was observed in NOESY spectra of D2−D3. Figure 4 shows overlapped NOEs in open bars; it can clearly be seen that there is a large amount of overlap present in the spectra of D1−D4, indicating the difficulties associated with the assignment of D1−D4 without the assignment information provided by the spectra of D2−D3.

Initial analysis of the NMR data for D1−D4 shows that the data reflect clearly the unfolded nature of the protein. For example, the large majority of the cross-peaks in the D2−D3 spectrum lie almost exactly in the same positions as cross-peaks for the corresponding residues in the D1−D4 spectrum, showing that the chemical shifts of the resonances in D2−D3 are determined by the local sequence and not by long-range effects. Cross-peaks for a small number of residues at the N- and C-termini of D2−D3 do not overlay with D1−D4 cross-peaks since the chemical shifts of these residues are affected by the presence of electrostatic charges at the chain termini of D2−D3. Similarly a comparison of the backbone ^{15}N chemical shifts and $^3J_{HN\alpha}$ coupling constants of residues from D1 (open symbols) and D2 (closed symbols) (Figure 5a,b) shows an excellent agreement for the chemical shifts and coupling constants of the amino acids that are the same in the two D repeats. For the purposes of comparison, 38 (the number of residues in D1) has been subtracted from the amino acid numbers for residues in D2 (for example, Asp-68 in D2 is equivalent to Asp-30 in D1 in this comparison).

For the $^3J_{HN\alpha}$ coupling constants, all but one of the values measured for D1−D4 and D2−D3 lie between the values predicted for residues in regular secondary structure, 4.8 Hz in α helices and 8.5 Hz in β strands (*38*). Except in the case of resonance overlap, all possible NN(i,i+1), αN(i,i+1), and scN(i,i+1) NOEs are observed throughout the sequence of D1−D4. These data are indicative of motional averaging showing that each residue is populating both α and β main-chain ϕ,ψ conformations. In addition, 20 NN(i,i+2), 59 αN(i,i+2), 39 scN(i,i+2), 3 αN(i,i+3), and 4 scN(i,i+3) NOEs have been observed; however no NOEs further apart than (i,i+3) in the sequence could be assigned in any NOESY spectra of either D1−D4 or D2−D3, indicating a lack of long-range persistent tertiary structure. Thus, the NMR data for D1−D4 are significantly different from those observed for a globular folded protein; indeed the parameters closely resemble those seen for other proteins in highly denatured states.

The close similarity of the NMR data for D1−D4 to data seen for denatured proteins prompted us to use similar approaches for analysis to those used with considerable success in studies of the latter. Unfolded proteins rapidly switch between many conformations, and in general the interconversion between conformers is fast, leading to an averaging of the observed NMR parameters (*47*, *48*). The interpretation of NMR data derived from non-native states is therefore complex due to the heterogeneity of the conformational ensemble adopted. However, some progress has been made, particularly in the case of chemical shifts, by comparing the experimental data with empirical values derived from experimental studies of short model peptide fragments (*45*, *49*, *50*). Another procedure that has proved particularly successful uses coupling constants and NOEs predicted from a statistical model for a random coil as a framework for interpreting the experimental data (*38*, *40*, *51*). Predictions from this statistical model are of considerable interest since they enable the experimental data to be interpreted in terms of specific torsion angle populations (*52*). In the case of D1−D4, both of these approaches have been used to help interpret the experimental NMR data.

The chemical shifts of D1−D4 and D2−D3 have been compared with values measured in unstructured model peptides; deviations from these empirical values for the backbone ^{15}N and H$^\alpha$ chemical shifts are shown in Figure 6. The deviations for ^{15}N chemical shifts take into account the identity of the preceding residue and were calculated from data from the peptide series GGXA (*45*). The value of the ^{15}N chemical shift of Ala-4 (A4) in these peptides shows marked differences depending on the identity of the third residue in the sequence (X3), which can be used to calibrate the deviation of the ^{15}N chemical shift of a particular residue depending on the identity of residue (i−1). The deviations for the H$^\alpha$ chemical shifts were calculated using data from

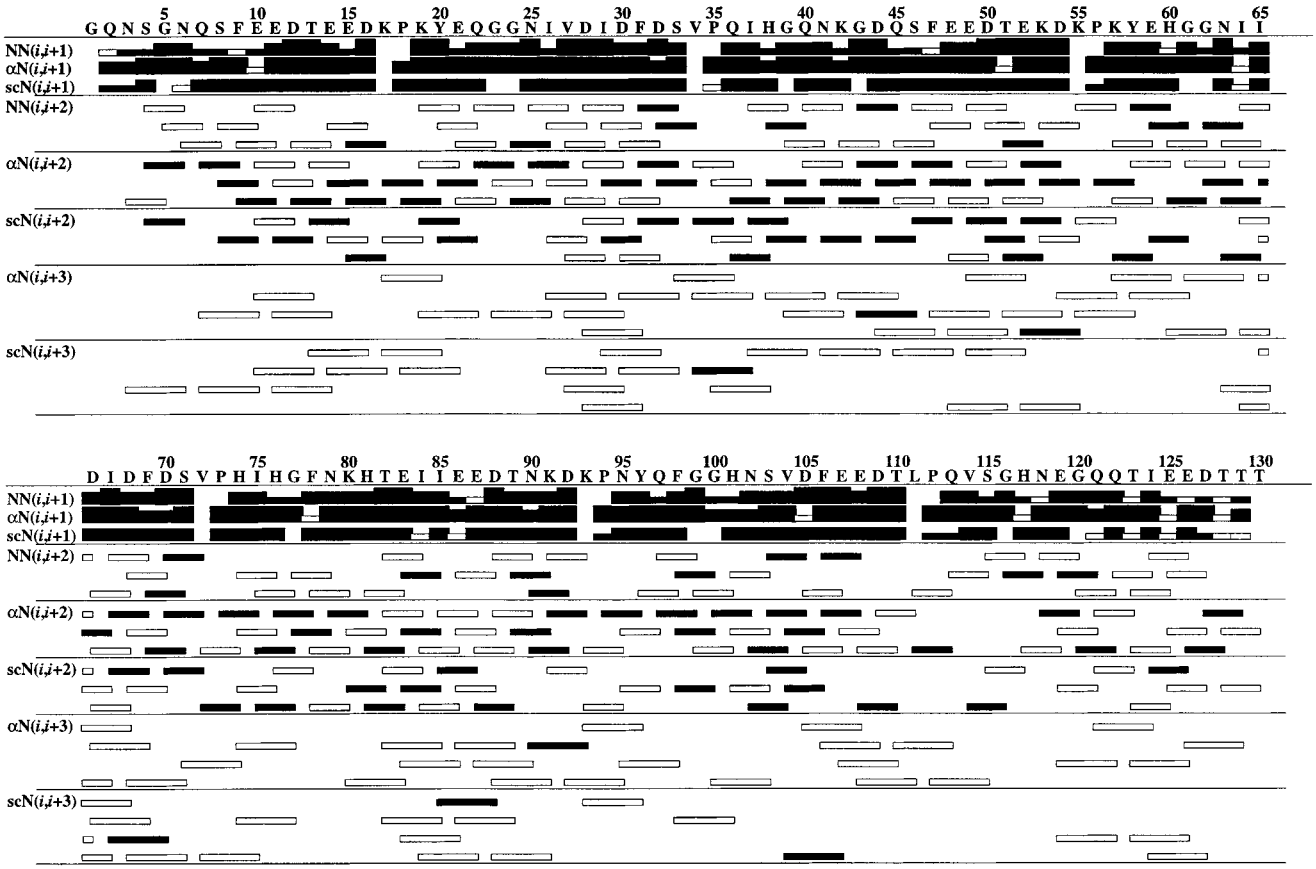

FIGURE 4: Summary of short- and medium-range NOEs for D1−D4 found in NOESY spectra recorded with mixing times of 200 ms. Closed bars represent observed NOEs, and open bars represent NOEs that cannot be observed due to spectral overlap.

the series of peptides GGXGG; these peptides are believed to minimize any structural preferences that may affect the highly conformation-dependent H^{α} chemical shifts (*49*). The ^{15}N chemical shifts for D1−D4 show an excellent overall agreement with the random coil values. The agreement for the H^{α} chemical shifts is less good, although two recent studies have also found differences in the H^{α} chemical shifts measured in non-native states of proteins with those measured in model peptides (*33*, *53*). Figure 6 clearly shows that the values of the H^{α} chemical shifts for the six residues that directly precede prolines are all shifted significantly downfield. This can be attributed to the fact that prolines cause the preceding residue to sample preferentially the β region of ϕ,ψ space (*54*, *55*).

Statistical Analysis. In the statistical model for a random coil, it is assumed that in this state there are no specific nonlocal interactions between residues in the polypeptide chain (*52*). Individual residues will however have local conformational preferences, and a description of the ϕ,ψ population of each residue in the protein is taken from the amino acid specific distributions of ϕ,ψ torsion angles in the protein database. For all residue types except glycine there is a clear preference for the α and β regions of ϕ,ψ space in the database, but the relative populations of these two regions differ significantly from one amino acid to another reflecting the characteristics of the side chains concerned (*37*, *38*, *51*). The model assumes that by taking all the structures in the database the effects of specific nonlocal interactions present in individual proteins will be averaged out. The protein database ϕ,ψ distributions were used to predict random coil coupling constants and NOE intensities for the D1−D4 sequence using the procedures described in the Materials and Methods section. The predictions were then compared with the experimental data.

As we have reported previously, good agreement is observed between the coupling constants predicted for the different types of amino acid in a random coil and the experimental $^3J_{HN\alpha}$ values for D1−D4 and D2−D3 [correlation coefficient 0.83; RMSD 0.25 Hz (*12*)]. Furthermore, we have identified from the coupling constant data recorded for D1−D4 and D2−D3 more subtle sequence-dependent variations in random coil ϕ,ψ populations superimposed on the overall amino acid-dependent variations (*12*). In particular, when the preceding residue has a β-branched or aromatic side chain, a significant increase occurs in the population of the less sterically restricted b region of ϕ,ψ space. Our previous work has shown that the $^3J_{HN\alpha}$ coupling constant data for D1−D4 give an even better agreement with predicted values from a random coil model that takes into account the effect of the residue at the ($i-1$) position [correlation coefficient 0.92; RMSD 0.22 Hz (*12*)].

An overall comparison of the experimental ($i,i+1$), ($i,i+2$), and ($i,i+3$) NOEs of D1−D4 and the predictions for a random coil is summarized in Table 1. NOEs that are predicted but could not be observed due to spectral overlap (shown in Figure 4) are excluded from all comparisons of experimental and predicted data. Sequential $\alpha N(i,i+1)$ and $NN(i,i+1)$ NOEs are predicted to be observed for all residues in D1−D4 (excluding proline) reflecting the significant population of both α and β conformers in a random coil;

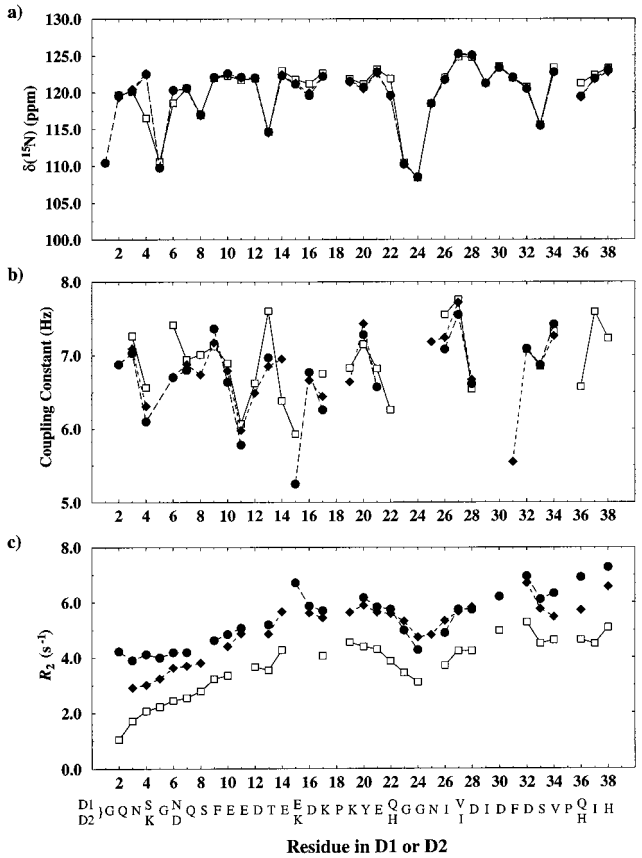

FIGURE 5: (a) Backbone ^{15}N chemical shifts, (b) experimental $^{3}J_{HN\alpha}$ coupling constants, and (c) backbone ^{15}N R_2 values (at a ^{1}H frequency of 600 MHz) of residues from D1 in D1−D4 (open squares), residues from D2 in D1−D4 (closed circles), and residues from D2 in D2−D3 (closed diamonds). The amino acid numbers used for residues from D2 have been obtained by subtracting 38 (the number of residues in D1) from the numbering in D1−D4.

this is indeed observed experimentally. A significant number of (i,i+2) NOEs is also predicted, particularly the αN(i,i+2) NOEs; the αH−NH(i,i+2) distance is short in mixed $\beta(i)\alpha(i+1)$ conformers that have a considerable population in a random coil. Excellent agreement is found for these αN(i,i+2) NOEs, 95% of the NOEs that are observed in NOESY spectra of D1−D4 being predicted by the random coil model, and 81% of the predicted NOEs being observed. For the other types of medium-range NOEs the agreement is somewhat less good, and the overall predictions show that 76% of observed (i,i+2) and (i,i+3) NOEs are predicted and that 67% of predicted NOEs can be observed experimentally.

Variations in the ϕ,ψ distributions of the different amino acids in the protein database result in variations in the predicted intensities of a given type of NOE for different amino acid sequences (*33*, *40*). In this work a cutoff was used based on the signal-to-noise levels in the spectra, all NOEs with an intensity greater than the cutoff being predicted to be observed. The intensity of certain medium-range NOEs such as the NN(i,i+2) and αN(i,i+3) NOEs for a random coil are very close to the cutoff used in the predictions for D1−D4. Consequently, even slight differences in the ϕ,ψ distributions between amino acids can result in a given NOE being included for one local sequence and excluded for another in the list of predicted random coil NOEs. In contrast to subtle variations in the amino acid distributions, large variations in signal-to-noise levels can be observed in experimental spectra resulting from line broadening effects such as differences in relaxation properties or solvent exchange rates. These are likely to be responsible for many of the discrepancies seen between the NOEs predicted for a random coil and those observed experimentally. To estimate the extent of variations in NOE intensity resulting from line broadening effects in the spectra of D1−D4, the intensity of αN(i,i+2) cross-peaks in two 3D NOESY−HSQC spectra, recorded at different spectrometer frequencies, were compared. Significant differences in intensity are seen between equivalent cross-peaks in the two spectra; this variation is larger than the range of intensities (relative values 0.4−0.9) predicted for random coil αN(i,i+2) NOEs for the D1−D4 sequence. Thus, intensity variations in the spectra are likely to be the major cause of the discrepancies between the experimental data for D1−D4 and the random coil predictions rather than deviations in the behavior of the protein from the random coil model.

Hydrodynamic Radius Measurements and ^{15}N Relaxation Data

Measurements. Analysis of chemical shift, coupling constant, and NOE data for an unfolded protein gives insight into the local conformational properties. For a full characterization, an understanding of the global properties of the system is also of importance. An unfolded protein will adopt an ensemble of conformations, but a measure of the average effective hydrodynamic radius of the ensemble can be obtained through diffusion coefficient measurements. These studies have been performed, as described in the Materials and Methods section, for both D1−D4 and D2−D3 in the presence of a small quantity of dioxan. The signal from dioxan decays much more rapidly than those from the protein (Figure 7); from the ratio of the observed diffusion constants for the protein and dioxan, values for the protein effective radius were calculated using a scaling factor determined previously from measurements on native folded hen lysozyme (*26*). By this procedure, the radii of D1−D4 (130 residues) and D2−D3 (77 residues) were determined to be 26.2 ± 0.1 and 18.5 ± 0.1 Å, respectively, D2−D3 being approximately 30% smaller than D1−D4. Comparison of the value for D1−D4 with the radii of gyration of native proteins containing a similar number of residues, such as ribonuclease A [124 residues, 15.0 ± 0.2 Å (*56*)], α-lactalbumin [123 residues, 15.5 Å (*57*)], and hen lysozyme [129 residues, 15.3 ± 0.2 Å (*58*)], shows that on average within the ensemble D1−D4 is almost 75% larger than would be expected for a compact fold. The value of the effective radius for D1−D4 is similar to that seen for denatured proteins with no disulfide bridges to restrict the conformational space accessible to the polypeptide chain such as reduced ribonuclease A [28.0 ± 1.0 Å (*56*)] and acid-unfolded cytochrome *c* [24.2 ± 0.3 Å (*59*)]. Therefore at both a local and a global level the conformational properties of D1−D4 resemble those of many proteins under strongly denaturing conditions.

Data from ^{15}N relaxation studies of D1−D4 and D2−D3 have been used to probe the dynamical properties of the protein. Relaxation data could be extracted for about 100 residues of D1−D4 and about 60 residues of D2−D3. The ^{15}N R_1, R_2 and {^{1}H}−^{15}N NOE values for D1−D4 and D2−D3 at a ^{1}H frequency of 600 MHz are plotted as a function

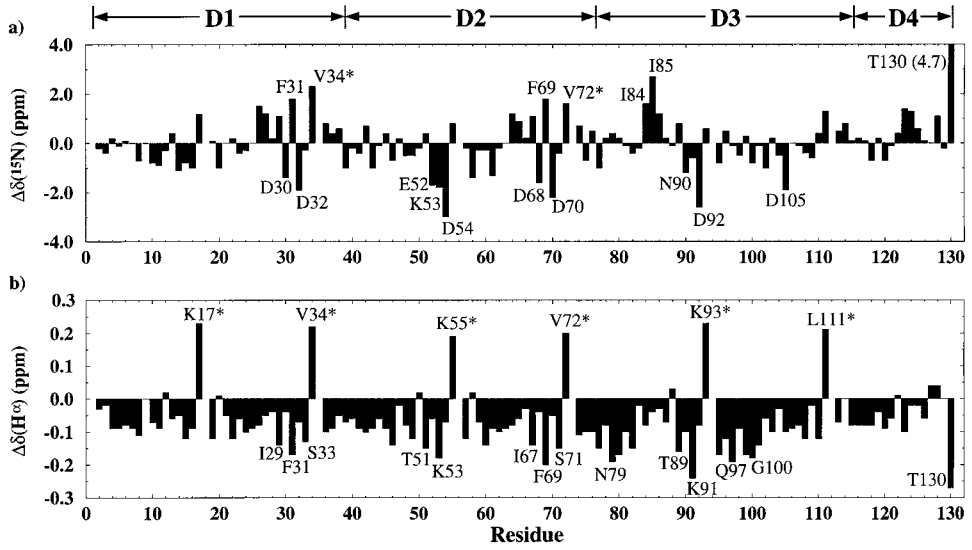

FIGURE 6: (a) ^{15}N and (b) H$^{\alpha}$ chemical shift deviations from empirical values measured in short unstructured peptides ($\Delta\delta = \delta^{\text{experimental}} - \delta^{\text{empirical}}$) for backbone resonances of D1−D4 at pH 6.0 and 5 °C. Sequence-corrected empirical ^{15}N chemical shifts were obtained from Braun et al. (*45*) H$^{\alpha}$ chemical shifts were taken from Merutka et al. (*49*). Residues of interest are labeled; residues labeled with asterisks directly precede proline residues.

Table 1: Overall Comparison for (*i*,*i*+1), (*i*,*i*+2) and (*i*,*i*+3) NOEs in D1−D4 between Experimental Data and Predictions for a Random Coil[a]

NOE	observed	predicted	O not P[b]	P not O[c]	O and P[d]
NN(*i*,*i*+1)	107	107	0	0	107
αN(*i*,*i*+1)	114	114	0	0	114
scN(*i*,*i*+1)	100	100	0	0	100
NN(*i*,*i*+2)	20	11	13	4	7
αN(*i*,*i*+2)	59	69	3	13	56
scN(*i*,*i*+2)	39	57	9	27	30
αN(*i*,*i*+3)	3	3	2	2	1
scN(*i*,*i*+3)	4	2	3	1	1
all[e]	125	142	30	47	95

[a] Predicted NOEs were calculated as described in Materials and Methods. Experimental NOEs were identified in 3D NOESY−HSQC spectra recorded at a ^{1}H frequency of 500 MHz with mixing times of 150 or 200 ms. A D1−D4 concentration of approximately 1.5 mM was used at pH 6 and 5 °C. [b] Observed but not predicted. [c] Predicted but not observed. [d] Observed and predicted. [e] All (*i*,*i*+2) and (*i*,*i*+3) NOEs.

of residue number in Figure 8. A more complete set of relaxation data recorded for D1−D4 and D2−D3 is available in the Supporting Information. It can be seen that the R_1 values are extremely uniform along the whole sequence of both D1−D4 and D2−D3 (apart from small differences at the chain termini); there is, however, much greater heterogeneity in the R_2 values with values varying by a factor of 3 between different regions of the chain in D1−D4 and by a factor of 2 for different regions in D2−D3. The possibility of chemical exchange effects being present in the R_2 values was checked by using transverse rotating frame ($R_{1\rho}$) experiments, which reduce the contributions from chemical exchange effects (on milli- to microsecond time scales) to the transverse relaxation rates (*60*). The $R_{1\rho}$ and R_2 values were found to be very similar for almost all residues, indicating the absence of significant chemical exchange contributions to the R_2 values. Figure 8 also shows that there is some heterogeneity in the heteronuclear NOE values of D1−D4 and D2−D3. The R_1, R_2 and heteronuclear NOE data all exhibit different values close to the N- and C-termini of the chains from the values seen in the central portion of the

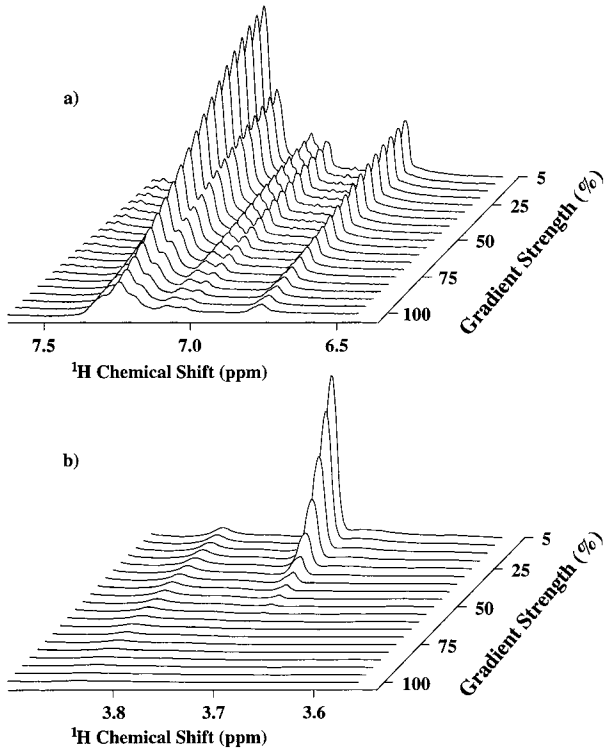

FIGURE 7: (a) Aromatic region and (b) expansion of the dioxan signal of the ^{1}H NMR spectra of D1−D4 obtained with the PG-SLED sequence at different pulsed field gradient strengths. The signal of the small reference molecule dioxan decays much faster than the signals from the protein. Decay rates are proportional to the diffusion coefficients for the protein and dioxan molecules, which are inversely proportional to their effective hydrodynamic radii.

chain, which is particularly conspicuous for the heteronuclear NOE data.

A direct comparison of ^{15}N R_1, R_2 and heteronuclear NOE values for both D1−D4 and D2−D3 shows that there is excellent agreement for all of the relaxation parameters measured in D1−D4 and D2−D3 between residues 50 and 110 of D2 and D3 (the central portion of D2−D3). This

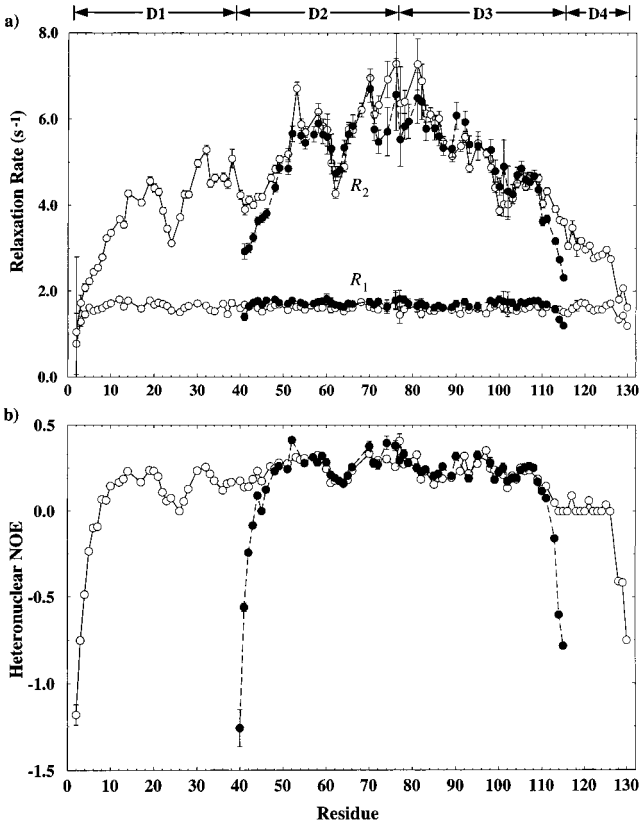

FIGURE 8: Comparison of (a) ^{15}N R_1, R_2 and (b) {^{1}H}−^{15}N NOE values for backbone amides from D1−D4 (open circles) and from D2−D3 (closed circles) at a ^{1}H frequency of 600 MHz.

strongly suggests that the dynamical processes present in both of these proteins are very similar except at the chain termini. This is a particularly significant result as the different measured effective radii of the two proteins mean that the overall rotational correlation times of D1−D4 and D2−D3 should differ by a factor of more than 3 (*61*). There would therefore be considerably different values for the R_2/R_1 ratios of equivalent residues in D1−D4 and D2−D3 if the dominant motional process giving rise to the relaxation of the ^{15}N nuclei was the overall rotation of the molecules. As this is not observed experimentally, local motions of the polypeptide chain must clearly dominate the relaxation process.

All of the relaxation parameters show differences between the corresponding residues in D1−D4 and D2−D3 at the N- and C-termini of D2−D3, which indicates that there is additional conformational freedom for these residues at the termini of D2−D3. This difference in relaxation properties for D1−D4 and D2−D3 in these regions, which have the same amino acid sequences, strongly suggests that there are different dynamical effects present at the termini relative to the central regions of unfolded polypeptide chains. The R_2 values for D1−D4 and D2−D3 differ for about 10 and 7 residues at the N- and C-termini of D2−D3, respectively. This difference in relaxation behavior at the termini together with the apparent lack of an overall correlation time for either D1−D4 or D2−D3 indicates that there is a persistence length of about 7 residues on either side of an amino acid that determines the dynamics of that amino acid in the unfolded protein. This persistence length is comparable to values of 10 residues determined by X-ray scattering techniques for denatured states of apocytochrome *c* and phosphoglycerate kinase and for the unfolded protein prothymosin α (*62*), and 7 residues determined by fitting ^{15}N relaxation data for denatured lysozyme to a model for a random coil (*33*).

Analysis. The relaxation data for D1−D4 have been compared with those from studies of other proteins. As the apparent lack of the dependence of the relaxation data on a global correlation time for D1−D4 and D2−D3 implies that an analysis of the relaxation data in terms of a Lipari−Szabo model-free method (*63*) is not appropriate, we have instead compared each of the different relaxation parameters separately on a semiquantitative basis with data in the literature. In these comparisons, it is important to note that variations in temperatures and solution viscosities in the different studies will affect the rates of the dynamical motions to some extent. Overall a close resemblance is observed to data seen in studies of denatured states rather than folded globular proteins. Thus, the values of the heteronuclear NOE for D1−D4 almost all fall in the range −0.2 to +0.4 (mean, +0.12 ± 0.24) similar to values observed in denatured proteins (*33*, *64*−*66*) and are significantly smaller than those observed for the native states of proteins with a similar size to D1−D4 including ribonuclease H (*67*), interleukin-4 (*68*), and lysozyme (*69*) (average NOE values greater than +0.6). Furthermore, the R_2 values for D1−D4 (largely ranging between 3 and 7 s^{-1}; mean 4.42 ± 1.36 s^{-1}) are smaller than those of native proteins such as staphylococcal nuclease (*30*), calmodulin (*34*), and lysozyme (*69*) (greater than 8 s^{-1} for most residues at 35 °C; the R_2 value is expected to be much greater at 5 °C for these proteins). They do however show similarities in both magnitude and heterogeneity to those observed in other studies of denatured proteins where values typically range from about 2 to 8 s^{-1} (*33*, *64*−*66*, *70*). The R_1 values for D1−D4 are also similar to those observed in denatured proteins, where little variation is observed around a mean value of about 1.6 s^{-1} (*33*, *64*−*66*, *70*).

For D1−D4 using a simple model of segmental motion, individual segments moving independently in solution, it would be expected that to a first approximation the observed R_2 values would be uniform along the sequence except at the termini. Although this is observed to a first approximation, as mentioned previously variations of up to a factor of 2 are seen (excluding residues at the termini). Analysis of this variation shows that in general the minima in the R_2 values are located around the glycine residues. This reflects the fact that glycine residues lack a C^β substituent and consequently populate much more of ϕ,ψ space than other residue types (including the α_L region) giving rise to a much greater degree of flexibility. Similar effects have been observed previously in partially folded proteins (*71*). The clearest minima are present around the conserved Gly-Gly motifs in D1 (Gly-23 and Gly-24) and D2 (Gly-61 and Gly-62) where the pairs of glycines will give even more flexibility to the main chain. Some of the maxima in the R_2 values occur where there are prolines in the sequence restricting backbone flexibility. Other maxima in the R_2 values may reflect the presence of persistent nonrandom features reducing flexibility such as hydrophobic clusters in the conformational ensemble for certain regions of the sequence. A number of studies have noted correlations between those regions with decreased mobility in unfolded states of proteins and

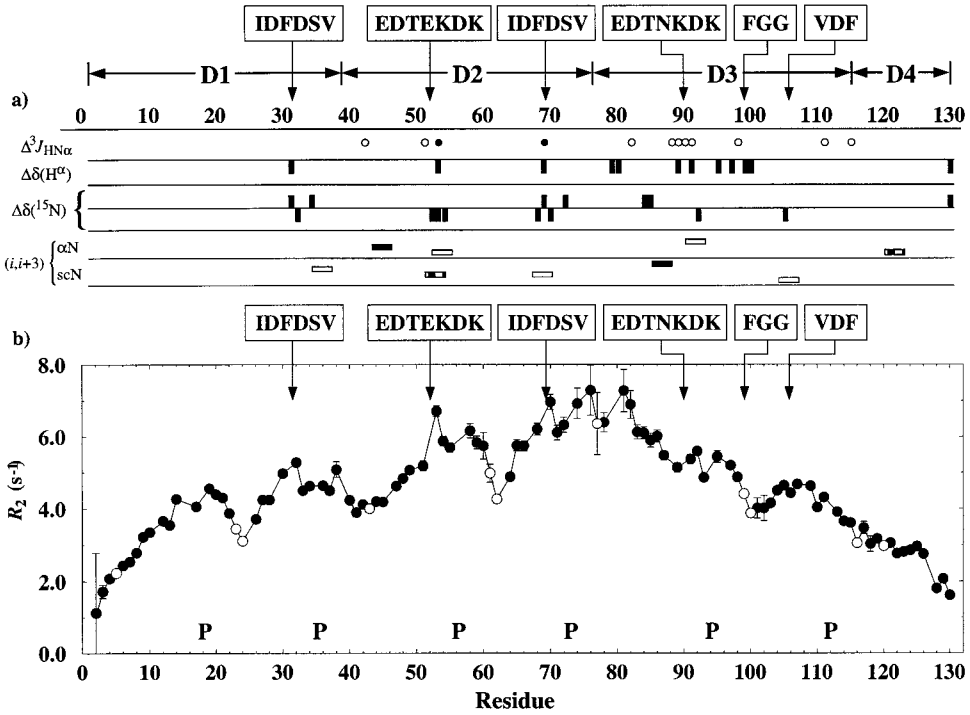

FIGURE 9: Summary of (a) structural and (b) dynamical NMR parameters measured for D1−D4 at pH 6.0 and 5 °C. (a) $^3J_{HN\alpha}$ coupling constants with marked deviations from the random coil coupling constant predictions that take into account the identity of the preceding residue [$\Delta^3J_{HN\alpha} = {}^3J_{HN\alpha}{}^{experimental} - {}^3J_{HN\alpha}{}^{predicted}$; (*12*)]: $\Delta^3J_{HN\alpha} < -0.4$ Hz are shown as open circles, and $\Delta^3J_{HN\alpha} < -0.8$ Hz are shown as closed circles. Chemical shifts with marked deviations from empirical values for unstructured peptides (see Figure 6) are shown as vertical bars (from top to bottom): $\Delta\delta(H^\alpha) < -0.15$ ppm (negative bars), $\Delta\delta(^{15}N) > 1.5$ ppm (positive bars) and $\Delta\delta(^{15}N) < -1.5$ ppm (negative bars). Comparison between predicted and experimental (*i*,*i*+3) NOEs; predicted NOEs were calculated as described in Materials and Methods. NOE cross-peaks that are observed and predicted are shown as closed bars, NOEs that are observed but not predicted are shown as open bars, and NOEs that are predicted but not observed are shown as dashed bars. (b) ^{15}N R_2 values at a ^1H frequency of 600 MHz for D1−D4. This figure also shows the regions that encompass the medium-range interactions discussed in the text and the locations of prolines (indicated by P) and glycines (indicated by open circles).

hydrophobic sequences (*33*, *64*, *66*). This is considered further in the next section.

Comparison of the R_2 values of residues of D1 (from D1−D4) and D2 (from both D1−D4 and D2−D3) is shown in Figure 5c. For the data for D2, end effects are apparent in the relaxation data of D2 from D2−D3 for about 10 residues; there is, however, excellent agreement for the R_2 values of residues 10−38 of D2 from D1−D4 and D2−D3. In the comparison of the R_2 values for D1 and D2, it might be expected that for the first 10 residues of D1 the data will be affected by the increased mobility found at the N-terminus of D1−D4 but that following this the relaxation data for residues 10−38 of D1 from D1−D4 should be comparable to the relaxation data for the corresponding residues in D2 from D1−D4. Figure 5c shows clearly that this is not the case, the R_2 values of D1 being approximately 1 s^{-1} lower than those for D2. This is in contrast to the close similarity in the chemical shift and coupling constants seen for D1 and D2 (Figure 5a,b). The differences between the relaxation data for D1 and D2 suggest that the dynamical behavior of a specific residue in an unfolded protein depends to some extent on longer-range effects (although not tertiary effects), unlike NMR parameters such as chemical shifts or coupling constants that depend primarily on the local amino acid sequence. Interestingly, however, the patterns of local minima and maxima in the R_2 values are very similar for D1 and D2; for example, clear minima can be seen around Gly-24 in D1 and Gly-62 in D2, which is due to a Gly-Gly motif giving rise to increased local mobility.

Sequence-Specific Conformational Properties. Analysis of the chemical shift, coupling constant, NOE, diffusion, and relaxation data for D1−D4 shows clearly that at both a local and a global level the protein is essentially unfolded; extended rather than compact conformers predominate in the conformational ensemble, and taken overall, variations in NMR conformational parameters along the polypeptide chain reflect the conformational preferences of the individual amino acids rather than any nonlocal persistent interactions. Detailed analysis shows however that for certain specific regions of the sequence there are deviations between the measured NMR parameters and those expected for a statistical random coil. These are of particular interest in the case of this protein, which is biologically active, as they could reflect the presence of persistent features that may be involved in the binding of the protein to fibronectin. These features are now analyzed in more detail in order to attempt to understand the structural preferences that may give rise to the observed NMR data.

In this section, we consider variations in all the NMR parameters measured, a summary of which is given in Figure 9. However, as the regions of the chain with persistent nonlocal interactions tend to have the largest relaxation rates and the broadest lines in experimental spectra (*33*, *64*), the measurement of accurate coupling constants and the identification of NOEs is difficult for these sequences. We have therefore found previously that regions with deviations from random behavior can be most readily identified from perturbations in chemical shift and/or heteronuclear relaxation

data (*33*). Three key areas of the sequence can be identified for D1−D4.

The first involves the sequence EDT(E/N)KDK, which is seen in both D2 (Glu-49 to Lys-55) and D3 (Glu-87 to Lys-93). Figure 6 shows that there are marked deviations between the observed chemical shifts and those seen in short model peptides for residues in these regions. In addition, αN(*i*,*i*+3) NOEs have been identified between Glu-52 and Lys-55 and between Asn-90 and Lys-93 in D1−D4; these NOEs are not predicted to be sufficiently intense to be observed in the experimental spectra of D1−D4 behaving as a random coil. The $^3J_{HN\alpha}$ coupling constants for Lys-53 (5.3 Hz) and Lys-91 (5.7 Hz) are also smaller than those measured for other lysine residues in D1−D4 and D2−D3 and are less than the value predicted for a random coil (sequence-dependent prediction is 6.5 Hz). These deviations are likely to reflect a persistent electrostatic interaction between glutamate and lysine side chains that have (*i*,*i*+4) separations in these sequences. It has been reported previously from studies of model peptides that when glutamate and lysine residues are in a sequence with either a (*i*,*i*+3) or (*i*,*i*+4) spacing then there is an interaction between the oppositely charged side chains that stabilizes a helix (*72*); the two side chains primarily hydrogen-bonding with one another so that the interaction is largely independent of pH (*73*). The presence of such a persistent nonlocal side-chain interaction in D1−D4 favoring the population of α relative to β conformations for the main chain of the residues involved would be consistent with the αN(*i*,*i*+3) NOEs identified and the small values of the $^3J_{HN\alpha}$ coupling constants for Lys-53 and Lys-91. Other evidence to support this interpretation for D1−D4 comes from a comparison of the NMR parameters for D2 with those for D1 (Figure 5). In D1 one of the lysine residues in this sequence in D2 (Lys-53) is substituted by a glutamate residue, this one substitution leading to a loss in the deviations of the coupling constants and chemical shifts from the behavior expected for a random coil.

A second deviation from the behavior expected for a random coil is identified from the significantly upfield-shifted H^N chemical shift of Gly-100 compared with the chemical shifts of all the other glycine residues in D1−D4 (for example, see Figure 2). This upfield shift probably arises from a persistent interaction between the aromatic ring of Phe-98 and the amide group of Gly-100. Aromatic−amide (*i*,*i*+2) interactions of this type have been characterized in a range of peptides *(74−76)* and denatured proteins *(33, 70, 77−79)*. The interactions are thought to be essentially electrostatic in character. The nonuniform charge distributions on aromatic rings enables them to act as hydrogen bond donors or acceptors. In this case, the interaction between the aromatic ring and the glycine amide presumably enables the hydrogen bonding capacity of the amide group to be at least partially fulfilled *(75, 80)*.

The third region involves the sequence IDFDSV, which is present in D1 (Ile-29 to Val-34) and D2 (Ile-67 to Val-72). Here again marked deviations in the H^α and ^{15}N chemical shifts are observed from those seen in the short peptides (Figure 6), and in addition both Ile-29 and Ile-67 contain upfield-shifted $C^{\gamma 2}H_3$ methyl groups. Furthermore, NOEs can be observed between the Ile-67 $C^{\gamma 2}H_3$ methyl group and the backbone amides of Phe-69 and Asp-70, although overlap precludes observation of similar NOEs for Ile-29. The $^3J_{HN\alpha}$ coupling constant of Phe-69 (5.6 Hz) is also much smaller than the value predicted for a random coil (7.0 Hz), indicating a preference for α rather than β main chain conformations for this residue. These NMR data are likely to reflect the presence of a hydrophobic cluster involving the isoleucine, valine, and phenylalanine residues in the sequence, with a close proximity between the isoleucine side chain and phenylalanine aromatic ring with an (*i*,*i*+2) separation; such hydrophobic clusters involving the side chains of hydrophobic and aromatic amino acids have been observed in a number of peptide fragments *(81, 82)* and denatured proteins (*33, 83*).

In the case of D1−D4, further insight into the preferred conformation for the isoleucine side chain and the phenylalanine ring comes from $^3J_{\alpha\beta}$ coupling constant values. For isoleucine residues, the $^3J_{\alpha\beta}$ coupling constant is predicted for a random coil to be 8.9 Hz *(84)*; this predicted value compares well with the experimental $^3J_{\alpha\beta}$ coupling constant obtained for Ile-84 of 8.5 Hz [measured in COSY spectra of D2−D3 using the method of Bartik and Redfield (*85*)] but is significantly different from the experimental value for Ile-67 of 6.5 Hz. This suggests that Ile-67 has specific χ_1 preferences to allow the $C^{\gamma 2}H_3$ group to interact (in preference to the CH_2CH_3 branch of the side chain) with the aromatic ring of Phe-69. In D1−D4 there is also evidence for an similar Val-Phe (*i*,*i*+2) interaction involving Val-104 and Phe-106. However, the chemical shifts and coupling constants for residues around the Val-Phe (*i*,*i*+2) interaction show smaller deviations from nonrandom values than for residues around the two Ile-Phe (*i*,*i*+2) interactions, perhaps because both $C^{\gamma}H_3$ groups of Val-104 are able to interact with the aromatic ring of Phe-106 giving rise to increased motional averaging.

In all of the regions considered here, in which persistent nonlocal interactions give rise to deviations in the conformational NMR parameters from those expected for a random coil, increased R_2 values are also observed compared with neighboring groups in the sequence. Thus, these persistent features appear to introduce an additional degree of stiffness into the polypeptide chain around the residues concerned. This is particularly clear when the R_2 values for D1 and D2 are compared since D1 lacks the Glu-Lys (*i*,*i*+4) interaction observed in D2 (Figure 5). Such conclusions explain previous results from studies of denatured lysozyme where increased R_2 values were seen in the vicinity of tryptophan side chains presumably involved in hydrophobic clusters (*33*).

In the light of the unusual structural and dynamical properties of D1−D4 for a biologically active protein, it is interesting to compare the conformational properties of the protein determined here with the results of applying secondary structure prediction algorithms to the D1−D4 sequence (Table 2). The Chou−Fasman (CF) (*41*) and Garnier−Osguthorpe−Robson (GOR) (*42*) statistical methods both predict helix in the N-terminal regions of each of the D repeats. In addition, the CF method predicts helix in the C-terminal part of D3. The PHD method (profile network system from Heidelberg) that uses multilayered neural networks incorporating evolutionary information (*43*, *44*), in contrast, produced very different results with no helix and only about 25% sheet being predicted for the whole of D1−D4; this was reduced to about 10% sheet when the sequences

Table 2: Secondary Structure Predictions for D1, D2, D3, and D4 Using CF (Chou−Fasman) (*41*), GOR (Garnier−Osguthorpe−Robson) (*42*), and PHD (Profile Network System from Heidelberg) (*43, 44*)[a]

D1	GQNSGNQSFEEDTEEDKPKYEQGGNIVDIDFDSVPQIH
CF	.TTTTTttHHHHHHHHHttt.TTtBBBBB.tt.BBBBB
GOR	TT........HHHHHHHHHHHTTT.BBBBBTT...TT.
PHD	LLLLLLLLLlLLLlLLLLllllLLlEEEElLLLLlee
D2	GQNKGDQSFEEDTEKDKPKYEHGGNIIDIDFDSVPHIH
CF	.TTTTtttHHHHHHHHHttt.tTTBBBBB.tt......
GOR	TT......HHHHHHHHHHHHHTTT.BBBBBTT......
PHD	elLLLLLLlllLLlLLLLllllLLlEEEEeLLLLLleE
D3	GFNKHTEIIEEDTNKDKPNYQFGGHNSVDFEEDTLPQVS
CF	.ttHHHHHHHHHTTTT.TTt.tt.tttHHHHHHBBBBBt
GOR	......HHHHHHHTTT.TTTTTTT......TTT......
PHD	EeLLleEEEElLLLLLLleeelLLLLllllLLLLLLL
D4	GHNEGQQTIEEDTTT
CF	tttTTHHHHHHHtt.
GOR	..........TTT..
PHD	LLLLLleEEEEeLLL

[a] H refers to helix, B or E refers to β-strand or extended structure, T refers to turn and L refers to loop. Capital letters mean that the prediction is of greater significance.

of D1, D2, and D3 were multiply aligned. These regions of extended structures can primarily be found in the C-terminal parts of D1, D2, and D3 where the CF and GOR methods predict β-sheet structure in D1, D2, and D3 (not GOR).

These structure predictions can be compared with the predicted populations of the α and β regions of ϕ,ψ space for the D1−D4 sequence from the statistical model for a random coil (Figure 3b). In the C-terminal regions of D1, D2, and D3 there are significant numbers of residues with a high propensity for populating β main chain conformations in a random coil (residues 26−38 of D1 and D2 both contain five isoleucine or valine residues) while in the N-terminal part of D1, D2, D3, and D4 aspartate and glutamate residues predominate with a high propensity for α conformers in a random coil. The effects of the different conformational preferences are reflected in the experimental data of D1−D4, for example, in lower values of $^3J_{HN\alpha}$ coupling constants in the N-terminal parts of the D repeats as compared to the C-terminal parts (seen clearly in Figure 3a where the individual values are averaged over 5-residue windows). Thus the two statistical secondary structure predictions give the correct ϕ,ψ tendencies but overestimate the amount of secondary structure present. A similar discrepancy between the amount of predicted and experimental secondary structure was also found for the unfolded protein prothymosin α (*62*). The PHD method gives a more accurate estimate of the amount of secondary structure present in D1−D4, especially when a multiple sequence alignment is used.

Biological Implications. The unfolded nature of D1−D4 has consequences for its biological activity of binding to the N-terminal domain of fibronectin. Studies of synthetic peptide fragments have allowed the position of the binding regions within the sequence of the D repeats to be probed (*86, 87*). In particular it has been found that a peptide encompassing residues 93−113 of D1−D4 (the C-terminal part of D3) has significant binding affinity for fibronectin, although removal of residues at the N- or C-terminal regions of this peptide results in greatly diminished affinity (*86*). Furthermore, studies of a range of peptide fragments from *Streptococcus dysgalactiae* fibronectin-binding protein have shown that the GG(X2−4)(I/V)DFXXD(S/T) sequence (corresponding to the C-terminal regions of D1, D2, and D3) is crucial for binding affinity (*87*). In this C-terminal region of the D repeat sequence, there are preferences for populating the β region of ϕ,ψ space (Figure 3b). If these are the key residues involved in binding, these extended conformations may play an important role in enabling a substantial number of sequentially related residues to be presented to fibronectin for highly specific binding. In addition, an interesting feature of the unfolded nature of D1−D4 is that it allows hydrophobic groups such as isoleucine, valine, and phenylalanine side chains in the IDF and VDF motifs to be accessible for binding rather than being buried in the core as is usually the case if they are located in a globular protein.

It is interesting to consider whether any of the nonlocal interactions identified in specific regions of the D1−D4 sequence in this work may have a biological role. A number of the fibronectin-binding repeat sequences from the various microbial fibronectin-binding proteins are found to have glutamate and lysine residues in their N-terminal halves with (i,i+3) or (i,i+4) spacings along the amino acid sequences of the fibronectin-binding repeats (*2*). As this feature results in nonlocal persistent electrostatic interactions between the glutamate and lysine side chains in D1−D4, such interactions may be important for the full fibronectin-binding affinity of these proteins to be realized. Furthermore, the scissile bonds of Lys-53 and Lys-91 in D1−D4 may be protected from proteolytic cleavage by steric hindrance when the side chains of these lysine residues are involved in nonlocal interactions with those of Glu-49 and Glu-87, respectively. Similarly, the presence of medium-range side chain to side chain hydrophobic interactions between isoleucine, valine, and phenylalanine residues in the IDFDSV sequences may protect them from proteolysis by enzymes specific for large hydrophobic residues in unfolded proteins. Additionally, in contrast to the large number of proteases specific for basic amino acids, there appears to be a lack of proteases directed against acidic residues, the V8 protease, which cuts at glutamate, being a rare exception (*88*). This lack of proteases directed against acidic groups in addition to the nonlocal persistent interactions discussed above would allow much of the polypeptide chain of D1−D4, with its very acidic nature, to remain largely protected from proteolytic cleavage.

CD studies have previously suggested that D1−D4 undergoes conformational changes on binding (*10*) although it remains largely unstructured. There is likely to be a substantial entropic penalty associated with the ordering the polypeptide chain of D1−D4 on complex formation. This will cause the binding interaction to be characterized by a weaker affinity for a given specificity (*12, 89−93*). Thus,

the specificities and affinities of individual interactions may be compromised with the advantage that the overall interaction is highly specific. In addition the unfolded nature of D1−D4 may be an important feature enabling avoidance of the immune system; in an ensemble of highly disordered molecules too few individual interactions may be made with antibody molecules for tight binding to be present overall. The significance of this is supported by previous experimental results that show that blocking antibodies produced against various forms of the D1−D4 immunogen have been found to be largely ineffective (*11*, *94*−*96*). In light of these advantages of having an unfolded conformation, it is interesting that a number of other proteins involved in protein−protein interactions such as microtubule-associated protein (*97*), tau protein (*98*), the cyclin-dependent kinase inhibitor $p21^{Waf1/Cip1/Sdi1}$ (*93*), the transcriptional activation domain of the herpes virus protein VP16 (*99*, *100*), the non-Aβ component precursor (*101*), and the inhibitor of flagellum specific sigma factor σ^{28} (*102*) have also been found to be essentially unfolded. The detailed insight into the structural and dynamical properties of D1−D4 provided by the work reported here therefore has significance for understanding the properties of all the members of this group of unfolded but biologically active proteins.

ACKNOWLEDGMENT

We thank Diane McBay, Danuta Mossakowska, and Josie Wallace of SmithKline Beecham Pharmaceuticals, Harlow, U.K., for the supply of ^{15}N-labeled D1−D4 and D2−D3 with support from Ian A. Critchley. We would also like to thank Harald Schwalbe and Klaus Fiebig for valuable discussions, Deborah Wilkins for help with the diffusion measurements, and Jonathan Boyd and Nick Soffe for technical assistance with the NMR facilities.

SUPPORTING INFORMATION AVAILABLE

Tables of ^{1}H and ^{15}N resonance assignments, heteronuclear relaxation rates (R_1, R_2), and heteronuclear NOEs determined for D1−D4 and D2−D3 at pH 6.0 and 5 °C (11 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Hay, E. D. (1991) in *Cell Biology of the Extracellular Matrix* (Hay, E. D., Ed.) Plenum Press, New York.
2. Patti, J. M., Allen, B. L., McGavin, M. J., and Höök, M. (1994) *Annu. Rev. Microbiol. 48*, 585−617.
3. Potts, J. R., and Campbell, I. D. (1994) *Curr. Opin. Cell Biol. 6*, 648−655.
4. Kuusela, P. (1978) *Nature 276*, 718−720.
5. Signäs, C., Raucci, G., Jönnson, K., Lindgren, P.-E., Anantharamaiah, G. M., Höök, M., and Lindberg, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 699−703.
6. Jönsson, K., Signäs, C., Müller, H.-P., and Lindberg, M. (1991) *Eur. J. Biochem. 202*, 1041−1048.
7. Burnham, M. K., Chopra, I., Critchley, I. A., and Knowles, D. J. C. (1994) Patent, International Publication No. W0 94/18327.
8. Kline, J. B., Xu, S., Bisno, A. L., and Collins, C. M. (1996) *Infect. Immun. 64*, 2122−2129.
9. Schneewind, O., Mihaylova-Petkov, D., and Model, P. (1993) *EMBO J. 12*, 4803−4811.
10. House-Pompeo, K., Xu, Y., Joh, D., Speziale, P., and Höök, M. (1996) *J. Biol. Chem. 271*, 1379−1384.
11. Speziale, P., Joh, D., Visai, L., Bozzini, S., House-Pompeo, K., Lindberg, M., and Höök, M. (1996) *J. Biol. Chem. 271*, 1371−1378.
12. Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E., Smith, R. A. G., Dobson, C. M., and Smith, L. J. (1997) *J. Mol. Biol. 274*, 152−159.
13. Lawlor, E. J., Elson, S. W., Holland, S., Cassels, R., Hodgson, J. E., Lloyd, M. D., Baldwin, J. E., and Schofield, C. J. (1994) *Tetrahedron 50*, 8737−8748.
14. Tabor, S. (1990) in *Current Protocols in Molecular Biology* (Ausubel, F. A., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., Eds.) Greene Publishing and Wiley-Interscience, New York.
15. Miekka, S. I., Ingram, K. C., and Menache, D. (1982) *Thromb. Res. 27*, 1−14.
16. Bodenhausen, G., and Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185−189.
17. Kay, L. E., Marion, D., and Bax, A. (1989) *J. Magn. Reson. 84*, 72−84.
18. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry 28*, 6150−6156.
19. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc. 111*, 1515−1517.
20. Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T., and Gronenborn, A. M. (1990) *Biochemistry 29*, 3542−3556.
21. Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B., and Feeney, J. (1990) *J. Magn. Reson. 90*, 420−425.
22. Ikura, M., Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 9020−9022.
23. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
24. Stejskal, E. O., and Tanner, J. E. (1965) *J. Chem. Phys. 42*, 288−292.
25. Gibbs, S. J., and Johnson C. S. (1991) *J. Magn. Reson. 93*, 395−402.
26. Jones, J. A., Wilkins, D. K., Smith, L. J., and Dobson, C. M. (1997) *J. Biomol. NMR 10*, 199−203.
27. Kay, L. E., Brooks, B., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc. 111*, 5488−5490.
28. Kay, L. E., and Bax, A. (1990) *J. Magn. Reson. 86*, 110−126.
29. Redfield, C., Smith, L. J., Boyd, J., Lawrence, G. M. P., Edwards, R. G., Smith, R. A. G., and Dobson, C. M. (1991) *Biochemistry 30*, 11029−11035.
30. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry 28*, 8972−8979.
31. Boyd, J., Hommel, U., and Campbell, I. D. (1990) *Chem. Phys. Lett. 175*, 477−482.
32. Palmer, A. G., Skelton, N. J., Chazin, W. J., Wright, P. E., and Rance, M. (1992) *Mol. Phys. 75*, 699−711.
33. Schwalbe, H., Fiebig, K., 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.
34. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) *Biochemistry 31*, 5269−5278.
35. Palmer, A. G., Rance, M., and Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371−4380.
36. Barkhuijsen, H., de Beer, R., and van Ormondt, D. (1986) *J. Magn. Reson. 67*, 371−375.
37. Swindells, M. B., MacArthur, M. W., and Thornton, J. M. (1995) *Nat. Struct. Biol. 2*, 596−603.
38. Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M., and Dobson, C. M. (1996) *J. Mol. Biol. 255*, 494−506.
39. Pardi, A., Billeter, M., and Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741−751.
40. Fiebig, K. M., Schwalbe, H., Buck, M., Smith, L. J., and Dobson, C. M. (1996) *J. Phys. Chem. 100*, 2661−2666.
41. Chou, P. Y., and Fasman, G. D. (1974) *Biochemistry 13*, 222−245.
42. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol. 120*, 97−120.
43. Rost, B., and Sander, C. (1993) *J. Mol. Biol. 232*, 584−599.

44. Rost, B., and Sander, C. (1994) *Proteins 19*, 55−72.
45. Braun, D., Wider, G., and Wüthrich, K. (1994) *J. Am. Chem. Soc. 116*, 8466−8469.
46. Alexandrescu, A. T., Abeygunawardana, C., and Shortle, D. (1994) *Biochemistry 33*, 1063−1072.
47. Dobson, C. M. (1992) *Curr. Opin. Struct. Biol. 2*, 6−12.
48. Shortle, D. R. (1996) *Curr. Opin. Struct. Biol. 6*, 24−30.
49. Merutka, G., Dyson, H. J., and Wright, P. E. (1995) *J. Biomol. NMR 5*, 14−24.
50. Plaxco, K. W., Morton, C. J., Grimshaw, S. B., Jones, J. A., Pitkeathly, M., Campbell, I. D., and Dobson, C. M. (1997) *J. Biomol. NMR 10*, 221−230.
51. Serrano, L. (1995) *J. Mol. Biol. 254*, 322−333.
52. Smith, L. J., Fiebig, K. M., Schwalbe, H., and Dobson, C. M. (1996) *Folding Des. 1*, R95−R106.
53. Zhang, O., and Forman-Kay, J. D. (1997) *Biochemistry 36*, 3959−3970.
54. Schimmel, P. R., and Flory, P. J. (1968) *J. Mol. Biol. 34*, 105−120.
55. Hurley, J. H., Mason, D. A., and Matthews, B. W. (1992) *Biopolymers 32*, 1443−1446.
56. Sosnick, T. R., and Trewhella, J. (1992) *Biochemistry 31*, 8329−8335.
57. Dolgikh, D. A., Abaturov, L. V., Bolotina, I. A., Brazhnikov, E. V., Bychkova, V. E., Gilmanshin, R. I., Lebedev, Y. O., Semisotnov, G. V., Tikopulo E. I., and Ptitsyn O. B. (1985) *Eur. Biophys. J. 13*, 109−121.
58. Chen, L., Hodgson, K. O., and Doniach, S. (1996) *J. Mol. Biol. 261*, 658−671.
59. Kataoka, M., Hagihara, Y., Mihari, K., and Goto, Y. (1993) *J. Mol. Biol. 229*, 591−596.
60. Szyperski, T., Luginbühl, P., Otting, G., Güntert, P., and Wüthrich, K. (1993) *J. Biomol. NMR 3*, 151−164.
61. Carrington, D., and McLachlan, A. D. (1967) *Introduction to Magnetic Resonance*, Harper and Row, New York.
62. Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H. R., Müller-Frohne, M., Zirwer, D., Czarnecki, J., and Damaschun, G. (1995) *Biochemistry 34*, 13211−13218.
63. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 10*, 4546−4559.
64. Alexandrescu, A. T., and Shortle, D. (1994) *J. Mol. Biol. 242*, 527−546.
65. Frank, M. K., Clore, G. M., and Gronenborn, A. M. (1995) *Protein Sci. 4*, 2605−2615.
66. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1997) *Biochemistry 36*, 2390−2402.
67. Powers, R., Clore, G. M., Stahl, S. J., Wingfield, P. T., and Gronenborn, A. (1992) *Biochemistry 31*, 9150−9157.
68. Redfield, C., Boyd, J., Smith, L. J., Smith, R. A. G., and Dobson, C. M. (1992) *Biochemistry 31*, 10431−10437.
69. Buck, M., Boyd, J., Redfield, C., MacKenzie, D. A., Jeenes, D. J., Archer, D. B., and Dobson, C. M. (1995) *Biochemistry 34*, 4041−4055.
70. Logan, T. M., Thériault, Y., and Fesik, S. W. (1994) *J. Mol. Biol. 236*, 637−648.
71. Buck, M., Schwalbe, H., and Dobson, C. M. (1996) *J. Mol. Biol. 257*, 669−683.
72. Marqusee, S., and Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8898−8902.
73. Scholtz, J. M., Qian, H., Robbins, V. H., and Baldwin, R. L. (1993) *Biochemistry 32*, 9668−9676.
74. Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., and Wright, P. E. (1992) *J. Mol. Biol. 226*, 795−817.
75. Kemmink, J., van Mierlo, C. P. M., Scheek, R. M., and Creighton, T. E. (1993) *J. Mol. Biol. 230*, 312−322.
76. Kemmink, J., and Creighton, T. E. (1995) *J. Mol. Biol. 245*, 251−260.
77. Arcus, V. L., Vuilleumier, S., Freund, S. M. V., Bycroft, M., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 305−321.
78. Pan, H., Barbar, E., Barany, G., and Woodward, C. (1995) *Biochemistry 34*, 13974−13981.
79. Schulman, B. A., Kim, P. S., Dobson, C. M., and Redfield, C. (1997) *Nat. Struct. Biol. 4*, 630−634.
80. Worth, G. A., and Wade, R. C. (1995) *J. Phys. Chem. 99*, 17473−17482.
81. Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A., and Wright, P. E. (1992) *J. Mol. Biol. 226*, 819−835.
82. Lumb, K. J., and Kim, P. S. (1994) *J. Mol. Biol. 236*, 412−420.
83. Neri, D., Billeter, M., Wider, G., and Wüthrich, K. (1992) *Science 257*, 1559−1563.
84. West, N. J., and Smith, L. J. (1998) *J. Mol. Biol. 280*, 867−977.
85. Bartik, K., and Redfield, C. (1993) *J. Biomol. NMR 3*, 415−428.
86. McGavin, M. J., Raucci, G., Gurusiddappa, S., and Höök, M. (1991) *J. Biol. Chem. 266*, 8343−8347.
87. McGavin, M. J., Gurusiddappa, S., Lindgren, P.-E., Lindberg, M., Raucci, G., and Höök, M. (1993) *J. Biol. Chem. 268*, 23946−23953.
88. Karlin, S. (1995) *Curr. Opin. Struct. Biol. 5*, 360−371.
89. Schultz, G. E. (1979) in *Molecular Mechanisms of Biological Recognition* (Balaban, M., Ed.) pp 79−84, Elsevier/North-Holland Biomedical Press, Amsterdam..
90. Searle, M. S., Williams, D. H., and Gerhard, U. (1992) *J. Am. Chem. Soc. 114*, 10697−10704.
91. Dobson, C. M. (1993) *Curr. Biol. 3*, 530−532.
92. Spolar, R. S., and Record, M. T. (1994) *Science 263*, 777−784.
93. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 11504−11509.
94. Ciborowski, P., Flock, J.-I., and Wadström, T. (1992) *J. Med. Microbiol. 37*, 376−381.
95. Rozalska, B., Sakata, N., and Wadström, T. (1994) *APMIS 102*, 112−118.
96. Sun, Q., Smith, G. M., Zahradka, C., and McGavin, M. J. (1997) *Infect. Immun. 65*, 537−543.
97. Hernández, M. A., Avila, J., and Andreu, J. M. (1986) *Eur. J. Biochem. 154*, 41−48.
98. Schweers, O., Schönbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994) *J. Biol. Chem. 269*, 24290−24297.
99. Shen, F., Triezenberg, S. J., Hensley, P., Porter, D., and Knutson, J. R. (1996) *J. Biol. Chem. 271*, 4819−4826.
100. Shen, F., Triezenberg, S. J., Hensley, P., Porter, D., and Knutson, J. R. (1996) *J. Biol. Chem. 271*, 4827−4837.
101. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T. (1996) *Biochemistry 35*, 13709−13715.
102. Daughdrill, G. W., Chadsey, M. S., Karlinsey, J. E., Hughes, K. T., and Dahlquist, F. W. (1997) *Nat. Struct. Biol. 4*, 285−291.

BI9814080