Biochemistry

 Volume 33, Number 46 November 22, 1994

Accelerated Publications

Secondary Structure of the ETS Domain Places Murine Ets-1 in the Superfamily of Winged Helix−Turn−Helix DNA-Binding Proteins†

Logan W. Donaldson,‡ Jeannine M. Petersen,§ Barbara J. Graves,§ and Lawrence P. McIntosh*,‡

The Department of Biochemistry and Molecular Biology and the Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z3, and The Department of Cellular, Viral and Molecular Biology, University of Utah School of Medicine, Salt Lake City, Utah 84132

Received August 31, 1994®

ABSTRACT: The members of the *ets* gene family of transcription factors are characterized by a conserved 85-residue DNA-binding region, termed the ETS domain, that lacks sequence homology to structurally characterized DNA-binding motifs. The secondary structure of the ETS domain of murine Ets-1 was determined on the basis of NMR chemical shifts, NOE and *J*-coupling constraints, amide hydrogen exchange, circular dichroism, and FT-IR spectroscopy. The ETS domain is composed of three α-helices (H) and four β-strands (S) arranged in the order H1-S1-S2-H2-H3-S3-S4. The four-stranded antiparallel β-sheet is the scaffold for a putative helix−turn−helix DNA recognition motif formed by helices 2 and 3. The 25 residues extending beyond the ETS domain to the native C-terminus of the truncated Ets-1 also contain a helical segment. On the basis of the similarity of this topology with that of catabolite activator protein (CAP), heat shock factor (HSF), and hepatocyte nuclear factor (HNF-3γ), we propose that *ets* proteins are members of the superfamily of winged helix−turn−helix DNA-binding proteins.

Interactions between transcriptional regulatory proteins and promoter elements play a key role in the control of gene expression. The basis for promoter recognition has been investigated by structural studies of a growing number of DNA-binding proteins. These studies, in combination with sequence comparisons, have grouped many regulatory proteins into families that use related protein motifs for DNA binding (Pabo & Sauer, 1992; Burley, 1994). Assignment of a transcription factor to a structural motif family provides a framework for understanding the specificity of its DNA recognition. One set of important transcriptional regulators that has not yet been placed in such a family are the *ets* proteins.

The *ets* proteins include more than 20 DNA-binding proteins that have been grouped on the basis of a highly conserved, 85-residue region, referred to as the ETS domain (Karim *et al.*, 1992). The ETS domain mediates the binding of *ets* proteins as monomers to 20-bp DNA sites. Homologs of many of the *ets* genes have been found in diverse organisms, ranging from *Drosophila* to humans. Recognition sites for *ets* proteins, which most often contain the core sequence 5′-GGA-3′, have been found in many cellular and viral promoters, suggesting a role for the *ets* genes in transformation, cell growth, and differentiation (Macleod *et al.*, 1992; Wasylyk *et al.*, 1993).

On the basis of several criteria, the *ets* proteins have been proposed to be a unique class of DNA−binding proteins and to constitute a possible new structural motif family. The

† This work was supported by grants from the National Cancer Institute of Canada, with funds from the Canadian Cancer Society (L.P.M.), and the National Institutes of Health (GM 38663 and CA 42014 to B.J.G. and the University of Utah Cancer Center).

* Author to whom correspondence should be addressed.
‡ University of British Columbia.
§ University of Utah School of Medicine.

® Abstract published in *Advance ACS Abstracts,* October 15, 1994.

0006-2960/94/0433-13509$04.50/0

ETS domain does not share significant sequence similarity to the well-defined DNA-binding motifs. Furthermore, the DNA recognition element for the ETS domain has no nucleotide similarity to other consensus binding sites. In addition, the ETS domain mediates a distinctive pattern of both major and minor groove DNA contacts that cover an unusually large region for monomeric binding (Nye *et al.*, 1992).

To identify the structural motif used for DNA recognition by the ETS domain, an N-terminal truncation mutant of murine Ets-1 was characterized by heteronuclear NMR, CD,[1] and FT-IR spectroscopy. In this report, we present the secondary structure of the ETS domain derived from the analysis of NMR chemical shift, NOE, *J* coupling, and amide hydrogen exchange data. On the basis of the observed organization of secondary structural elements, *ets* proteins are proposed to be part of the superfamily of winged helix−turn−helix DNA-binding proteins, which includes the *Escherichia coli* catabolite activator protein (CAP) (Schultz *et al.*, 1991), *E. coli* biotin repressor (birR) (Wilson *et al.*, 1992), hepatocyte nuclear factor (HNF-3γ) (Clark *et al.*, 1993), yeast and *Drosophila* heat shock factor (HSF) (Harrison *et al.*, 1994; Vuister *et al.*, 1994a,b), and histone (H5) (Ramakrishnan *et al.*, 1993).

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification. A polypeptide containing the 110 C-terminal residues (amino acids 331−440) of murine Ets-1 was expressed in *E. coli* Topp-2 cells (Stratagene, La Jolla, CA) under the control of the tac promoter in the plasmid pCW (Gegner *et al.*, 1992). With the exception of the N-terminal methionine, no additional residues were added to the protein, which is denoted Ets-1ΔN331. Uniformly ^{15}N-labeled protein was produced from cells cultured in M9 minimal medium containing 1.0 g/L 99% $^{15}NH_4Cl$ or $^{15}NH_4SO_4$ and 1.0 g/L 99% ^{15}N-labeled Isogro algal extract (Isotec, Miamisburg, OH). Uniformly ^{15}N- and ^{13}C-enriched protein was produced from cells grown in M9 minimal medium containing 1.0 g/L 99% $^{15}NH_4Cl$ and 2.0 g/L 99% [$^{13}C_6$]glucose supplemented with 1.0 g/L 99% ^{15}N/^{13}C-labeled algal extract. The amides of Tyr, Lys, Val, and Leu were selectively ^{15}N-labeled in Topp-2 cells using the protocol described by McIntosh and Dahlquist (1990).

Ets-1ΔN331 was purified by FPLC on anion and cation exchange columns (Petersen *et al.*, manuscript in preparation). The final yield of Ets-1ΔN331 was ~3.5 mg/L bacterial culture, using a calculated $\epsilon_{280} = 3.2 \times 10^4$ M^{-1} cm^{-1} in 6 M guanidinium hydrochloride (Edelhoch, 1967). The Ets-1ΔN331 was estimated to be ~95% pure by SDS−polyacrylamide electrophoresis. Electrophoretic mobility shift assays showed that Ets-1ΔN331 retained sequence specific, high-affinity binding to an Ets-1 consensus site (Petersen *et al.*, manuscript in preparation). The first five residues of Ets-1ΔN331 were confirmed as Gly-Ser-Gly-Pro-Ile by Edman sequencing.

CD and FT-IR Spectroscopy. The circular dichroism spectrum of 15 μM Ets-1ΔN331 (20 mM sodium phosphate and 0.5 M NaCl (pH 6.4, 25 °C)) in a 0.1-cm jacketed quartz cell was acquired with a Jasco J-730 CD spectropolarimeter. The spectrum was solvent-subtracted and processed using Jasco software. The Fourier transform infrared (FT-IR) spectrum of a 6-μL sample of 1.0 mM Ets-1ΔN331 (20 mM sodium phosphate, 500 mM NaCl, and 10 mM DTT (pH 6.5)) in a 6-μm path length calcium fluoride cell was acquired on a Perkin-Elmer Model 2000 FT-IR spectrometer with a spectral range of 2200−1200 cm^{-1} and a spectral resolution of 2 cm^{-1}. The spectrum was processed according to the method of Dong *et al.* (1990).

NMR Spectroscopy. Experiments were performed on a Varian Unity 500 NMR spectrometer equipped with three radio-frequency channels and a pulsed-field-gradient accessory. The NMR data were processed and analyzed using Felix v2.30 software (Biosym Technologies, San Diego, CA). Identical buffer conditions (20 mM sodium phosphate (pH 6.45), 0.5 M sodium chloride, 10 mM DTT, 0.01% sodium azide, and 10% D_2O or 99% D_2O, 20 °C) were used for all Ets-1ΔN331 samples.

The following experiments were recorded with a 0.7 mM sample of uniformly ^{15}N-labeled Ets-1ΔN331: 2D HSQC (Kay *et al.*, 1992) and HMQC-J (Kay & Bax, 1990) and 3D TOCSY-HSQC, NOESY-HSQC (Fesik & Zuiderweg, 1990), and HNHA (Vuister & Bax, 1993). To monitor the time course of the amide hydrogen exchange, a series of HSQC spectra was acquired after transferring the sample into D_2O sample buffer (pH* 6.5) by passage through a Sephadex G-25 spin-column. After complete exchange, 2D homonuclear DQF-COSY, clean TOCSY (65-ms MLEV-17 mixing period), and NOESY (100-ms mixing period) spectra were recorded on the sample. ^{1}H−^{15}N HSQC spectra were also obtained from 0.2−0.4 mM samples of Ets-1ΔN331 selectively labeled with [α-^{15}N]tyrosine, -valine, -leucine, or -lysine.

The following 3D experiments were recorded with a 0.7 mM sample of uniformly ^{13}C/^{15}N-labeled Ets-1ΔN331: HNCO (Muhandiram & Kay, 1994), HNCACB (Wittekind & Müeller, 1993), CBCA(CO)NH (Grzesiek & Bax, 1992), CBCACOHA (Kay, 1993), H(CCO)NH, C(CO)NH (Grzesiek *et al.*, 1993), and HCCH-TOCSY (Kay *et al.*, 1993). For experiments in which magnetization was detected on the amide H^N, sensitivity-enhanced gradient pulse sequences were employed (Muhandiram & Kay, 1994), along with selective "flip-back" pulses to return water magnetization to equilibrium before the acquisition of the NMR signal (Grzesiek & Bax, 1993).

RESULTS

Ets-1ΔN331 is an amino-terminal deletion mutant of the murine Ets-1 protooncoprotein, containing the ETS domain (residues 331−415) and the native carboxyl terminus (residues 416−440). The latter 25 amino acids were found to be essential for solubility. Ets-1ΔN331 represents a minimal-sized fragment of Ets-1 that is fully functional for DNA binding and amenable for structural studies by NMR methods.

[1] Abbreviations: CD, circular dichroism; DQF-COSY, double quantum filtered correlation spectroscopy; DTT, dithiothreitol; ets, E26 transformation specific; Ets-1ΔN331, the N-terminal deletion fragment of murine Ets-1 consisting of residues 331−440; FT-IR, Fourier transform infrared; HSQC, heteronuclear single quantum correlation; wHTH, winged helix−turn−helix; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; pH*, the observed pH meter reading without correction for isotope effects; TOCSY, total correlation spectroscopy.

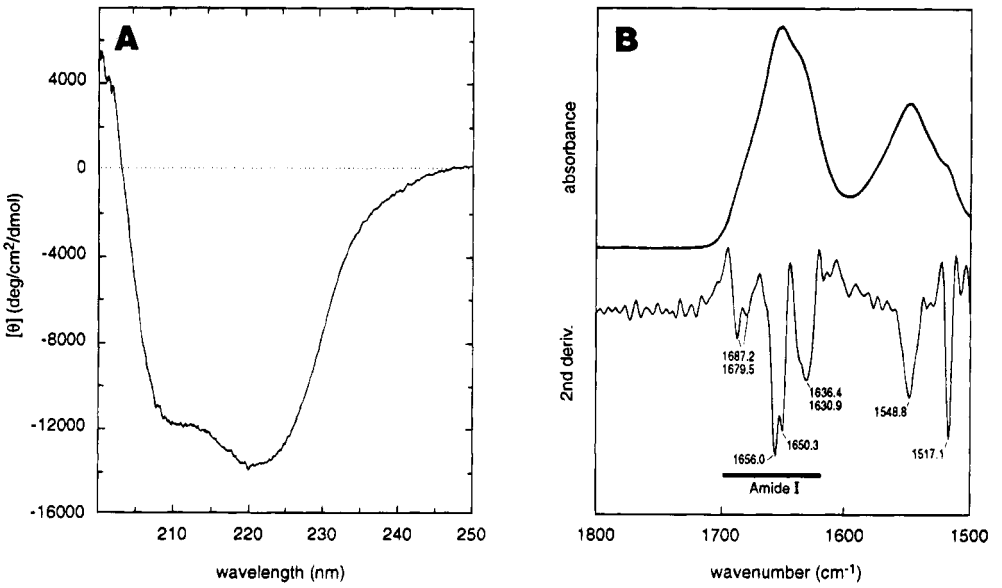

FIGURE 1: CD (A) and FT-IR (B) spectra of Ets-1ΔN331 demonstrating the presence of α-helical and β-sheet conformations. The far-UV CD spectrum of Ets-1ΔN331 shows distinct minima at 210 and 222 nm that are characteristic of α-helices. In the second derivative (thin line) of the amide I region of the FT-IR transmission spectrum (bold line), the peaks at (1687.2 cm^{-1}, 1679.5 cm^{-1}), 1656.0 cm^{-1}, 1650.3 cm^{-1}, and (1636.4 cm^{-1}, 1630.9 cm^{-1}) arise from β-turn, α-helix, extended, and β-sheet conformations, respectively (Surewicz *et al.*, 1993). A second, primarily α-helical contribution (1548.8 cm^{-1}) lies in the amide II region of the spectrum. The peak at 1517.1 cm^{-1} is characteristic of tyrosyl vibrations.

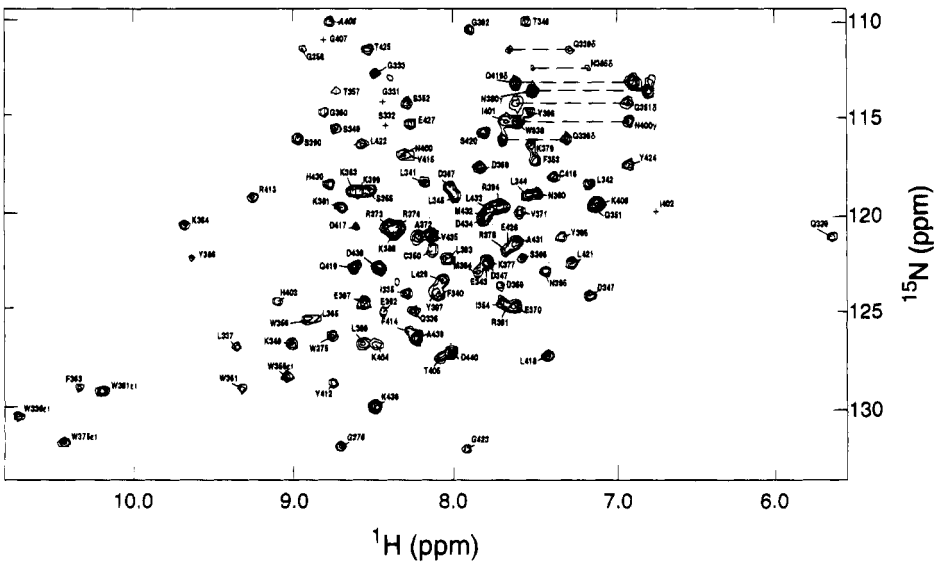

FIGURE 2: $^{1}H-^{15}N$ HSQC spectrum of Ets-1ΔN331 at pH 6.45 and 20 °C. Weak peaks from unusually broadened resonances and from the amides of a minor population of the protein retaining the N-terminal methionine are indicated by a +. This N-terminal heterogeneity results from incomplete processing of the amino terminus in the preparation of labeled protein.

Circular Dichroism and FT-IR Spectroscopy. The global secondary structure of Ets-1ΔN331 was characterized initially using CD and FT-IR spectroscopy. These studies demonstrated that under solution conditions similar to those used for the subsequent NMR studies, the Ets-1 fragment is folded and contains both α-helices and β-strands. The CD spectrum of Ets-1ΔN331 shows distinct minima at 210 and 222 nm, characteristic of α-helical structures (Figure 1A). The set of minima observed in the second derivative of the amide I region of the FT-IR spectra of the protein is diagnostic of α-helical and, most importantly, β-sheet conformations (Figure 1B).

NMR Spectroscopy. The $^{1}H-^{15}N$ HSQC spectrum of uniformly ^{15}N-enriched Ets-1ΔN331 is essentially a complete fingerprint of the protein, with 101 of the expected 105 backbone amides detected (Figure 2). Near-complete assignments of the main-chain ^{1}H, ^{13}C, and ^{15}N resonances of Ets-1ΔN331 were obtained using a combination of three complementary approaches. First, the amide resonances of the tyrosine, valine, leucine, and lysine residues, which compose 32% of the protein sequence, were identified from $^{1}H-^{15}N$ HSQC spectra of selectively labeled Ets-1ΔN331 samples (not shown). These served as unambiguous reference points for the specific resonance assignments. Second, a combination of 3D $^{1}H-^{13}C-^{15}N$ scalar correlation experiments was used to derive intraresidue and sequential connectivities between the main-chain atoms in a sample of uniformly $^{15}N/^{13}C$-enriched Ets-1ΔN331 (Ikura *et al.*, 1990). Third, in cases where degeneracies or insufficient signal-to-noise ratios precluded confident assignments using these through-bond, triple-resonance experiments, a main-chain-directed strategy was followed to interpret the 3D $^{1}H-^{15}N$

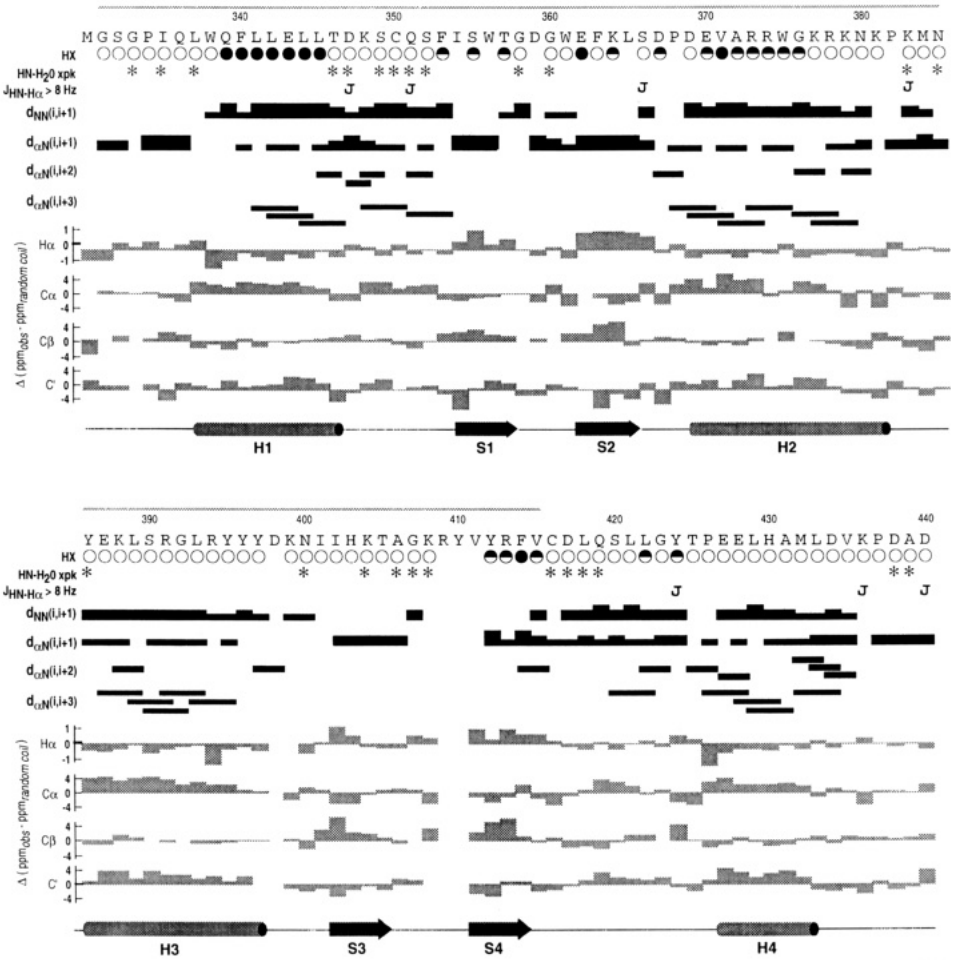

FIGURE 3: Summary of the NMR evidence used to derive the secondary structure of Ets-1ΔN331. The locations of the four α-helices (H) and four β-strands (S) are marked by cylinders and arrows, respectively. A bar above the protein sequence highlights the 85-residue ETS domain. (i) HX: Amide hydrogen−deuterium exchange measured after transfer of the protein to D_2O buffer at pH* 6.5 and 20 °C. Filled circles indicate slow amide exchange ($t_{1/2}$ > 1000 min), half-filled circles indicate intermediate exchange (50 < $t_{1/2}$ < 1000 min), and open circles indicate that a rate could not be determined under these conditions ($t_{1/2}$ < 50 min). (ii) H^N-H_2O xpk: An asterisk denotes rapid amide hydrogen exchange ($t_{1/2}$ < 0.5 s) or an NOE to long-lived bound water, as evidenced by a strong H^N-H_2O cross-peak in the NOESY-HSQC spectrum. (iii) $^3J_{HN-H\alpha}$: Amides with $^3J_{HN-H\alpha}$ > 8 Hz are marked by the letter J. (iv) Unambiguous NOEs between H^N and H^α are labeled as $d_{NN(i,i+1)}$, $d_{\alpha N(i,i+1)}$, $d_{\alpha N(i,i+2)}$, and $d_{\alpha N(i,i+3)}$. The relative NOE strength (weak/medium/strong) is reflected by the bar thickness (t_{mix} = 100 ms). (v) The main-chain H^α, C^α, C^β, and C′ chemical shifts are plotted as the difference from the corresponding random coil values (Wishart *et al.*, 1992; Wishart & Sykes, 1994).

TOCSY-HSQC and NOESY-HSQC spectra of uniformly ^{15}N-enriched Ets-1ΔN331 (Englander & Wand, 1987). Representative spectra of Ets-1ΔN331 and a table of the main-chain resonance assignments are presented in the supplementary material.

Secondary Structure Analysis by NMR. The secondary structure of Ets-1ΔN331 was derived using a combination of the following four NMR approaches (Figure 3).

(i) Chemical Shifts. A strong correlation exists between secondary structure and the deviations of the H^α, C^α, C^β and C′ chemical shifts of an amino acid in folded protein from its random coil values (Wishart *et al.*, 1992; Wishart & Sykes, 1994). For example, residues in α-helices show downfield or positive C^α and C′ secondary shifts, whereas those in β-sheets show positive H^α and C^β shift perturbations. The observed secondary shifts for Ets-1ΔN331 define four pronounced α-helical regions (H1−H4), as well as four short sequences with β-strand conformations (S1−S4).

(ii) Sequential and Short-Range NOE Correlations. Regular α-helical and β-sheet structures show characteristic NOE patterns between H^N, H^α, and H^β protons (Wüthrich, 1986). The NOE correlations between these main-chain protons in Ets-1ΔN331, obtained from 2D homonuclear NOESY and 3D heteronuclear $^1H-^{15}N$ NOESY-HSOC spectra, are summarized in Figure 3. Evidence for four α-helical regions in Ets-1ΔN331 is provided by a contiguous series of $H^N_i-H^N_{i+1}$ NOE interactions, combined with weak $H^\alpha_i-H^N_{i+1}$ NOEs and a limited number of observable $H^\alpha_i-H^N_{i+3}$ NOEs. The four stretches of residues in β-strand conformations are identified by relatively strong $H^\alpha_i-H^N_{i+1}$ NOEs and the absence of sequential $H^N_i-H^N_{i+1}$ NOEs. More diagnostically, the patterns of cross-strand NOE interactions between the H^α and H^N protons of these residues provide unambiguous evidence that Ets-1ΔN331 contains a four-stranded antiparallel β-sheet (Figure 4).

(iii) $^3J_{HN-H\alpha}$ Scalar Coupling. The $^3J_{HN-H\alpha}$ coupling constant is diagnostic of the conformation of a residue because of its dependence upon the ϕ backbone dihedral angle. Regular helical regions are characterized by $^3J_{HN-H\alpha}$ couplings on the order of 4−6 Hz, while peptides in extended conformations show couplings larger than 7 Hz (Wüthrich, 1986). Only $^3J_{HN-H\alpha}$ values larger than ~8 Hz were measured confidently due to the broad amide ^{15}N line widths (>12 Hz) in the HMQC-J spectrum of Ets-1ΔN331 and the

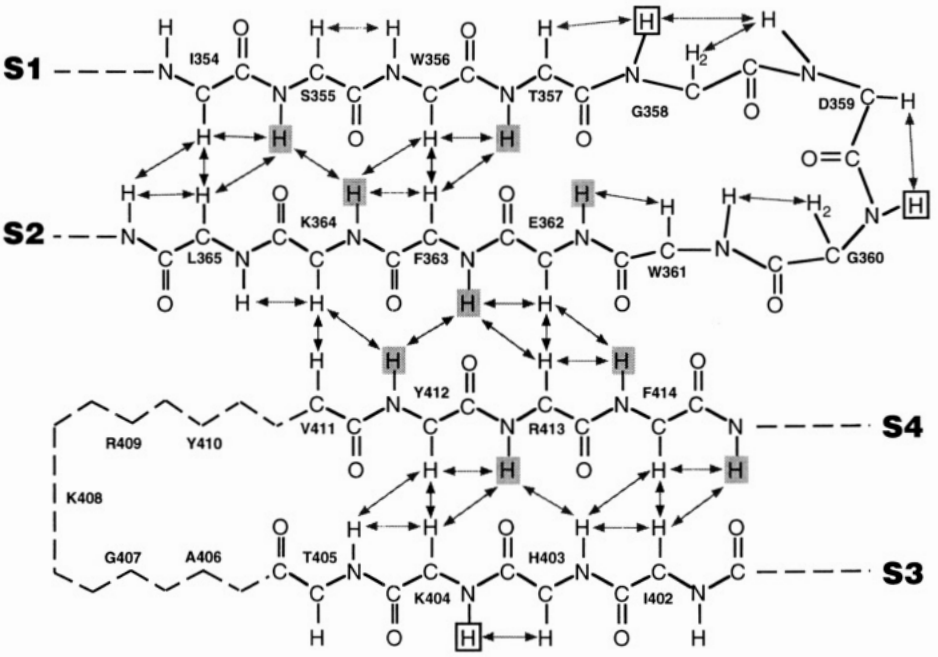

FIGURE 4: Schematic representation of the four-stranded antiparallel β-sheet in Ets-1ΔN331. The observed NOEs between H^N and H^α are indicated by arrows. The amide protons with intermediate or slow hydrogen−deuterium exchange kinetics are indicated by shaded boxed, whereas those giving strong exchange cross-peaks to H_2O in a NOESY-HSQC spectrum are in open boxes (see Figure 3). The stereochemistry of the hairpin turn has not been determined, and hydrogen bonds are not explicitly implied in this figure.

low signal-to-noise ratios in the HNHA spectrum (Figure 3). Consistent with the conclusions drawn from chemical shifts and NOE patterns, all of the residues assigned to the four helices showed couplings below this limit. The residues with $^3J_{HN-H\alpha}$ couplings greater than ~8 Hz were generally located near the ends of the helical regions.

(iv) Amide Hydrogen Exchange. The rate of the exchange of an amide H^N with water depends upon hydrogen bonding and solvent accessibility. Thirty amides in Ets-1ΔN331 with exchange rates of less than 2×10^{-4} s^{-1} were detected by recording a series of $^1H-^{15}N$ HSQC spectra after transfer of the protein into D_2O buffer at pH* 6.5. This corresponds to a protection of at least 10^4-fold relative to an amide in an unstructured polypeptide (Englander & Poulsen, 1969). The amides with significantly retarded exchange rates were located in helices H1 and H2, β-strands S1, S2, and S4, and a small section of the C-terminal sequence (Figures 3 and 4). This provides additional support for the hydrogen bonding expected within these secondary structural elements. Although the amides in helices H3 and H4 and strand S3 were not measurably protected under these conditions, the first HSQC spectrum was recorded 1 h after transfer of the protein to D_2O, and thus only those amide protons with significantly retarded (>10^4-fold) exchange kinetics were detected by this approach.

Twenty-four amides in Ets-1ΔN331 with fast exchange rates, comparable to those of an unstructured hydrated polypeptide ($t_{1/2}$ < 0.5 s), were identified from strong cross-peaks to water in a 3D $^1H-^{15}N$ NOESY-HSQC spectrum. These amides lie in loops and turns between the α-helices and β-strands (Figures 3 and 4). No attempt was made to distinguish exchange cross-peaks from NOEs to long-lived bound waters at this stage of the analysis.

The Secondary Structure of Ets-1ΔN331. Analysis of NMR chemical shift, NOE, *J* coupling, and amide hydrogen exchange data demonstrates that the first 85 residues of Ets-1ΔN331, corresponding to the ETS domain proper, have the secondary structure H1-S1-S2-H2-H3-S3-S4. The 25 residues beyond the ETS domain contain a fourth helical region, H4. The NMR evidence for this secondary structure is discussed here.

Helix H1 contains Leu337−Thr346, as evidenced by $H^N_i-H^N_{i+1}$ and $H^\alpha_i-H^N_{i+3}$ NOEs, the absence of $^3J_{HN-H\alpha} > 8$ Hz, negative H^α secondary shifts, and positive C^α and C' secondary shifts. Residues 339−345 also show strong protection from hydrogen−deuterium exchange, indicating that they constitute a stable hydrogen-bonded helix. Although sequential $H^N_i-H^N_{i+1}$ NOEs and positive C^α secondary shifts suggest that the helix may continue beyond Thr346, the large $^3J_{HN-H\alpha}$ measured for Asp347 and the fast amide exchange of residues 348−352 indicate that the backbone of these residues may form an irregular or flexible structure.

Helix H1 precedes a β-hairpin comprising strands S1 (Ile354−Thr357) and S2 (Glu362−Leu365) linked by a four-residue turn (Sibanda *et al.*, 1989). The β-sheet conformation of these amino acids is readily identified by strong $H^\alpha_i-H^N_{i+1}$ NOEs, distinct positive H^α and C^β secondary shifts, slow amide hydrogen exchange, and most importantly, cross-strand NOEs between residues in S1 and S2. The antiparallel alignment of the two strands is established from the $H^\alpha-H^\alpha$ NOE correlations between Ile354 and Leu365 and those between Trp356 and Phe363 (Figure 4). Presently, the turn cannot be unequivocally classified according to observed NOE patterns (Wüthrich, 1986).

Following the β-hairpin, Asp369 or possibly Pro368 initiates helix H2. This helix ends near residues 379−381, which appear to deviate from a standard helical conformation, as indicated by a relative loss of protection from amide hydrogen exchange, strong $H^\alpha_i-H^N_{i+1}$ NOEs, and irregular patterns of secondary chemical shifts. Helix H3 is composed of Tyr386−Tyr397, as identified by characteristic C^α and C' secondary shifts and $H^N_i-H^N_{i+1}$ and $H^\alpha_i-H^N_{i+3}$ NOEs. However, the protection from amide hydrogen exchange is less than that observed with helices H1 and H2. The residues

between helices H2 and H3, which include Lys383−Asn385, clearly are nonhelical, with parameters such as the large $^3J_{HN-H\alpha}$ coupling measured for Lys383 and fast amide hydrogen exchange. The exact size of this region, which posibly forms a turn between helices 2 and 3, remains ill-defined because the end of helix H2 is uncertain.

A second β-hairpin, formed by strands S3 (Ile402−Thr405) and S4 (Val411−Phe414) completes the secondary structure of the ETS domain. These two extended strands are defined primarily by pronounced downfield H^α and C^β secondary shifts and diagnostic cross-strand NOE connections. The amides in the five-residue turn or loop linking S3 and S4 are not protected from hydrogen exchange and appear to be exposed to solvent, giving exchange or possibly NOE correlations to H_2O in a $^1H-^{15}N$ NOESY-HSQC spectrum. Together, the two β-hairpins, S1−S2 and S3−S4, form a four-stranded antiparallel β-sheet with a distinct hydrophobic face (Figure 4). The alignment of the strands is (+1, +2*x*, −1) by the nomenclature of Richardson (1981), with strand S4 sandwiched between S2 and S3.

In the region bounded by residues Tyr397−Val411, many H^N and ^{15}N resonances were unusually weak or broad, thereby making assignments difficult to obtain. Asp398, Arg409, and Tyr410 could not be assigned, and combinations of triple-resonance experiments and selective isotopic labeling yielded only partial assignments for Lys408 and Val411. These residues include the region joining H3 and S3, strand S3 itself, and the turn or loop extending to S4. Line broadening due to slow conformational averaging of this region of Ets-1ΔN331 is the likely reason for the weak or missing resonances. The water magnetization is only minimally perturbed in the experiments employing pulsed-field gradients and selective flip-back pulses, and thus rapid amide hydrogen exchange is not expected to cause significant line broadening under these experimental conditions. It is noteworthy that Vuister *et al.* (1994a) reported similar behavior for the H^N and ^{15}N resonances from the corresponding residues in *Drosophila* HSF.

The final 25 amino acids in Ets-1ΔN331, although not part of the conserved ETS domain, also adopt a defined structure. With the exception of the extreme C-terminal residues Asp438, Ala439, and Asp440, which have unusually sharp resonances due to rapid conformational averaging, the H^N and ^{15}N peaks from this portion of Ets-1ΔN331 are comparable in line width and intensity to those from the bulk of the ETS domain. Helix H4 spans Glu427−Met432, using the criteria of contiguous $H^N_i-H^N_{i+1}$ NOEs and positive C^α and C' and negative H^α secondary chemical shifts. The residues immediately preceding Glu427 also show $H^N_i-H^N_{i+1}$ NOEs and protection from hydrogen exchange. However, the patterns of $H^\alpha_i-H^N_i$ NOEs, backbone *J* couplings, and secondary chemical shifts preclude a confident assignment of the secondary structure for this region of the protein.

DISCUSSION

Overview of the Secondary Structure of Ets-1ΔN331. NMR analysis demonstrates that the ETS domain is composed of three α-helices and a four-stranded antiparallel β-sheet. The three helices (H1−H3) and the four β-strands (S1−S4) are arrayed linearly in the order H1-S1-S2-H2-H3-S3-S4. Additional evidence supports this analysis: (i) Both α-helical and β-sheet secondary structural elements are observed through a combination of CD and FT-IR studies of Ets-1ΔN331. (ii) Given all *ets* protein sequences, the PHD algorithm (Rost & Sander, 1993) predicts a consensus secondary structure for the ETS domain closely resembling that determined by NMR methods (data not shown). (iii) In the alignment of over 20 ETS domains, amino acid insertions or deletions occur only between the helices and strands: H1 and S1, S1 and S2, S2 and H2, H2 and H3 (not shown).

The secondary structure analysis also indicates a fourth helical region in the C-terminal sequence following the ETS domain of Ets-1ΔN331. Although not formally considered part of the ETS domain, there is limited primary sequence conservation in this region among a subset of *ets* proteins (Klämbt *et al.*, 1993). This extension is also required to maintain the solubility of the murine Ets-1 ETS domain. Therefore, it is likely that this C-terminal extension is an important structural component of Ets-1. Several studies have in fact implicated this region in the regulation of Ets-1 DNA binding (Hagman *et al.*, 1992; Lim *et al.*, 1992).

The Winged Helix−Turn−Helix Superfamily. The ordering of α-helices and β-strands of the Ets-1 ETS domain strongly resembles that of the DNA-binding domains of *E. coli* catabolite activator protein (CAP), *E. coli* biotin repressor (birR), hepatocyte nuclear factor (HNF-3γ), yeast and *Drosophila* heat shock factors (HSF), and the globular domain of histone (H5) (Figure 5). The DNA-binding domains of these proteins also have similar tertiary structures (Brennan, 1993; Burley, 1994). Crystallographic and NMR analyses demonstrate that three helices in each of these domains pack to form a hydrophobic core, which rests on an amphiphatic β-sheet scaffold. DNA contacts are made through a helix−turn−helix motif (helices 2 and 3), and β-strands, the intervening loop regions, and the first helix. One particularly large loop within the β-sheet of HNF-3γ led to the designation winged helix to describe the structural motif of this protein (Clark *et al.*, 1993). We suggest the nomenclature winged helix−turn−helix (wHTH) for the entire superfamily of DNA-binding proteins. In this new, broader context, the wing would be the β-sheet region and any associated loops. This proposal distinguishes the wHTH superfamily from other HTH proteins, such as the homeodomain family, and highlights the role of the sheet region (with or without loops) in potential DNA contacts.

On the basis of the close resemblance of their secondary structures, we propose that the ETS domain has a tertiary structure similar to that of the wHTH proteins, with helices 2 and 3 corresponding to a helix−turn−helix motif. The structural similarity of the wHTH proteins occurs in the absence of any pronounced sequence similarity. Thus, previous attempts to align the ETS domain with the wHTH motif were largely unsuccessful due to the lack of clues derived from only primary sequence information.

Conserved Hydrophobic Core. Additional support for the classification of the ETS domain as a wHTH motif comes from the presence of conserved hydrophobic residues. The interiors of the wHTH proteins are formed by the packing of hydrophobic side chains from both the helices and the antiparallel β-sheet (Schultz *et al.*, 1991; Clark *et al.*, 1993; Harrison *et al.*, 1994; Vuister *et al.*, 1994b). The ETS domain has three amphipathic α-helices and a four-stranded antiparallel β-sheet with a distinct nonpolar face (Figures 4 and 5). Given the ordering of the α-helices with respect to

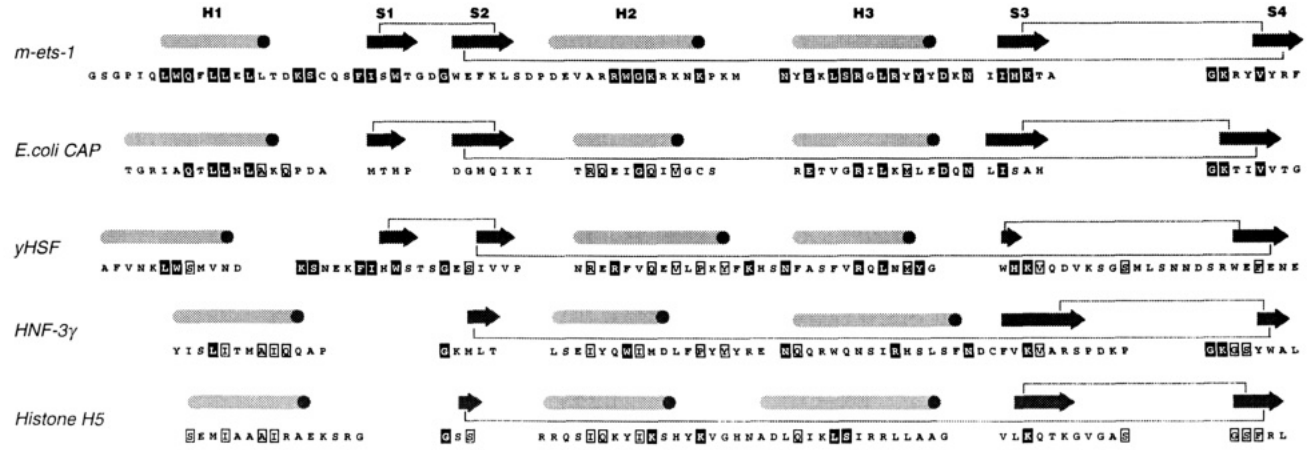

FIGURE 5: Alignment of the secondary structure determined for the ETS domain of Ets-1ΔN331 with the secondary structural elements observed in the crystal structures of the DNA-binding domains of CAP (Schultz *et al.*, 1991), yeast HSF (Harrison *et al.*, 1994), HNF-3γ (Clark *et al.*, 1993), and histone H5 (Ramakrishnan *et al.*, 1993). The α-helices (H) are shown as cylinders and the β-strands (S) as arrows. The topology of the β-sheet is marked by lines connecting the arrows. The second and third helices shown for each protein constitute the HTH motif. The sequences were aligned manually by helix and sheet elements and then translated to maximize possible sequence similarities. Gaps were introduced to accommodate the variable-sized turn or loop regions from the different proteins. A shaded box indicates a sequence identity between Ets-1 and any of the other four proteins, while an open box highlights an identical amino acid found in two or more of CAP, yeast HSF, HNF-3γ, and H5. *E. coli* biotin repressor (Wilson *et al.*, 1992; not shown) may also be aligned by its secondary structure to the proteins depicted here.

the β-sheet, it is likely that the ETS domain will also adopt this tertiary architecture.

The hydrophobic interior of the ETS domain is predicted to include the three highly conserved tryptophan residues Trp338, -356, and -375. Remarkably, the 18−19-residue spacing between these aromatic groups is similar to that observed in the *myb* DNA-binding proteins, thus prompting the suggestion that the *ets* and *myb* DNA-binding domains may be structurally similar (Anton & Frampton, 1988). An NMR analysis of the all-helical *myb* domain reveals that the conserved tryptophans, which lie in or immediately before three α-helices, pack to form a hydrophobic core (Ogata *et al.*, 1992). In contrast, the three tryptophans of the ETS domain, which are likely to fulfill a similar structural role, are located in two helices (H1, H2) and a β-strand (S1). Based on the secondary structure alignment in Figure 5, tryptophans are also preserved in H1 and S1 of the wHTH protein yeast HSF, while a phenylalanine is a conservative substitution for the tryptophan in H2. These three aromatic groups contribute to the hydrophobic core of HSF (Harrison *et al.*, 1993).

Mode of DNA Binding. DNA-binding studies also provide support for the classification of the ETS domain as a wHTH motif. Diverse DNA sequences are recognized by different wHTH proteins, as might be expected from the lack of amino acid conservation. However, several subgroups display distinct core recognition sequences: HSF, AGAA (Fernandes *et al.*, 1994); HNF-3 and *forked head* related proteins, TTTG (Zaret, personal communication); and the *ets* proteins, GGA-(A/T) (Karim *et al.*, 1992). Histone (H5) lends added diversity by binding DNA nonspecifically. In addition, the subunit configurations of superfamily members in the DNA-bound state are extremely diverse. HNF-3γ, H5, and Ets-1 bind as monomers, CAP and birR bind as dimers, and HSF binds as a trimer.

In spite of these diverse features, DNA−protein complexes of two wHTH proteins, CAP and HNF-3γ, display a striking concordance in the pattern of DNA contacts. As with other helix−turn−helix motifs (formed by helices 2 and 3), helix 3 binds the DNA in the major groove, while helix 2 stabilizes the position of this so-called recognition helix by making phosphate contacts. The β-sheet wing provides additional backbone contacts on the same face of the DNA that is contacted by helix 3. A C-terminal extension in HNF-3γ forms another extended structure to provide additional contacts, again on the same face of the DNA as the HTH motif. A function similar to the HNF-3γ carboxyl terminus is supported by helix 1 of each subunit of CAP. In summary, the interface with DNA is centered over a region of the major groove and extends over the minor groove in both flanking directions.

This mode of DNA binding is consistent with both genetic and biochemical data collected from Ets-1 DNA-binding studies. Methylation and ethylation interference analyses have demonstrated a central major groove contact zone over at least 5 bp that could be bound by helix 3, possibly in conjunction with helix 2 (Nye *et al.*, 1992). Helix 3 is very highly conserved among all ETS domains and could account for the strong common preference of GGA contacts in the major groove. An altered specificity mutation in the related ETS domain protein, Elf-1, maps to helix 3 (Bosselut *et al.*, 1994). This is consistent with the proposed role of this helix in DNA recognition and the description of helices 2 and 3 of the ETS domain as a helix−turn−helix motif. Hydroxyl radical and DNAse I footprinting data have mapped flanking regions of contact that span the minor groove on either side of the central major groove-binding region (Nye *et al.*, 1992). On the basis of the proposed wHTH tertiary structure of the Ets-1 ETS domain, the β-sheet wing could account for one set of minor groove flanking contacts, while the structured region in the carboxyl-terminal extension could contact the other flanking minor groove. Thus, the winged helix−turn−helix motif provides a framework to begin to understand the challenging problem of DNA-binding specificity of ETS domain proteins.

ACKNOWLEDGMENT

Dr. L. E. Kay is acknowledged foremost for providing pulse sequences and invaluable advice on triple-resonance

NMR spectroscopy. Dr. A. G. Mauk provided access to CD and FT-IR equipment, and Dr. D. Thackray provided expert assistance. We also acknowledge Dr. T. Alber, Dr. J. Skalicky, and Dr. Peter Flynn for helpful discussions, L. Cuddeford for assistance with DNA-binding studies, and S. Adu for inspiration throughout this study.

SUPPLEMENTARY MATERIAL AVAILABLE

One table listing the main-chain H^N, ^{15}N, H^α, $^{13}C^\alpha$, $^{13}C^\beta$, and $^{13}C'$ resonances of Ets-1ΔN331 and one figure illustrating CBCA(CO)NH/HNCACB and HSQC-TOCSY/NOESY correlations (6 pages). Ordering information is given on any current masthead page.

REFERENCES

Anton, I. A., & Frampton, J. (1988) *Nature 336*, 719.
Bosselut, R., Levin, J., Adjadj, E., & Ghysdael, J. (1993) *Nucleic Acids Res. 21*, 5184.
Brennan, R. G. (1993) *Cell 74*, 773.
Burley, S. K. (1994) *Curr. Opin. Struct. Biol. 4*, 3.
Clark, K. L., Halay, E. D., Lai, E., & Burley, S. K. (1993) *Nature 364*, 412.
Dong, A., Huang, P., & Caughey, W. S. (1990) *Biochemistry 29*, 3303.
Edelhoch, H. (1967) *Biochemistry 6*, 1948.
Englander, S. W., & Poulsen, A. (1969) *Biopolymers 7*, 379.
Englander, S. W., & Wand, A. J. (1987) *Biochemistry 26*, 5953.
Fernandes, M., Xiao, H., & Lis, J. T. (1994) *Nucleic Acids Res. 22*, 167.
Fesik, S. W., & Zuiderweg, E. R. (1990) *Q. Rev. Biophys. 23*, 97.
Gegner, J. A., Graham, D. R., Roth, A. F., & Dahlquist, F. W. (1992) *Cell 70*, 975.
Grzesiek, S., & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291.
Grzesiek, S., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 12593.
Grzesiek, S., Anglister, J., & Bax, A. (1993) *J. Magn. Reson. Ser. B 101*, 114.
Hagman, J., & Grosschedl, R. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 8859.
Harrison, C. J., Bohm, A. A., & Nelson, H. C. M. (1994) *Science 263*, 224.
Ikura, M., Kay, L. E., & Bax, A. (1990) *Biochemistry 29*, 4659.
Karim, F. D., Urness, L. D., Thummel, C. S., Klemsz, M. J., McKercher, S. R., Celada, A., Beveren, C. V., Maki, R. A., Gunther, C. V., Nye, J. A., & Graves, B. J. (1990) *Genes Dev. 4*, 1451.
Kay, L. E. (1993) *J. Am. Chem. Soc. 115*, 2055.
Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110.
Kay, L. E., Keifer, P., & Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663.
Kay, L. E., Xu, G.-Y., Singer, A. U., Muhandiram, D. R., & Forman-Kay, J. D. (1993) *J. Magn. Reson. 101*, 333.
Klämbt, C. (1993) *Development 117*, 163.
Lim, F., Kraut, N., Frampton, J., & Graf, T. (1992) *EMBO J. 11*, 643.
Macleod, K., Leprince, D., & Stehelin, D. (1992) *Trends Biochem. Sci. 17*, 251.
McIntosh, L. P., & Dahlquist, F. W. (1990) *Q. Rev. Biophys. 23*, 1.
Muhandiram, D. R., & Kay, L. E. (1994) *J. Mag. Reson. Ser. B 103*, 203.
Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D., & Graves, B. J. (1992) *Genes Dev. 6*, 975.
Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., & Nishimura, Y. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6428.
Pabo, C. O., & Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053.
Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., & Sweet, R. M. (1993) *Nature 362*, 219.
Richardson, J. (1981) *Adv. Protein Chem. 34*, 167.
Rost, B., & Sander, C. (1993) *J. Mol. Biol. 232*, 584.
Schultz, S. C., Shields, G. C., & Steitz, T. A. (1991) *Science 253*, 1001.
Sibanda, B. L., Blundell, T. L., & Thornton, J. M. (1989) *J. Mol. Biol. 206*, 759.
Surewicz, W. K., Mantsch, H. H., & Chapman, D. (1993) *Biochemistry 32*, 389.
Vuister, G. W., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772.
Vuister, G. W., Kim, S.-J., Wu, C., & Bax, A. (1994a) *Biochemistry 33*, 10.
Vuister, G. W., Kim, S.-J., Orosz, A., Marquardt, J., Wu, C., & Bax, A. (1994b) *Nature Struct. Biol. 9*, 605.
Wasylyk, B., Hahn, S. L., & Giovane, A. (1993) *Eur. J. Biochem. 211*, 7.
Wilson, K. P., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J., & Matthews, B. W. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9257.
Wishart, D. S., & Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171.
Wishart, D. S., Richards, F. M., & Sykes, B. D. (1992) *Biochemistry 31*, 1647.
Wittekind, M. G., & Mueller, L. (1993) *J. Magn. Reson. Ser. B 101*, 201.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.