Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K., & Gronenborn, A. M. (1990) *Biochemistry 29*, 9324–9334.
Parraga, G., Horvath, S., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., & Klevit, R. E. (1988) *Science 241*, 1489–1492.
Parraga, G., Horvath, S., Young, E. T., & Klevit, R. E. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 137–141.
Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Weiss, M. A., & Keutmann, H. (1990) *Biochemistry 29*, 9808–9813.
Weiss, M. A., Mason, K. A., Dahl, C. E., & Keutmann, H. (1990) *Biochemistry 29*, 5660–5664.
Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Alternating Zinc Fingers in the Human Male Associated Protein ZFY: 2D NMR Structure of an Even Finger and Implications for "Jumping-Linker" DNA Recognition†

Michel Kochoyan,‡,§ Timothy F. Havel,‡ Dzung T. Nguyen,‖ Charles E. Dahl,‡ Henry T. Keutmann,‖ and Michael A. Weiss*,‡,‖

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, and Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114

Received October 17, 1990; Revised Manuscript Received December 26, 1990

ABSTRACT: ZFY, a sex-related Zn-finger protein encoded by the human Y chromosome, is distinguished from the general class of Zn-finger proteins by the presence of a two-finger repeat. Whereas odd-numbered domains and linkers fit a general consensus, even-numbered domains and linkers exhibit systematic differences. Because this alternation may have fundamental implications for the mechanism of protein–DNA recognition, we have undertaken biochemical and structural studies of fragments of ZFY. We describe here the solution structure of a representative nonconsensus (even-numbered) Zn finger based on 2D NMR studies of a 30-residue peptide. Structural modeling by distance geometry and simulated annealing (DG/SA) demonstrates that this peptide folds as a miniglobular domain containing a C-terminal β-hairpin and N-terminal α-helix ($\beta\beta\alpha$ motif). These features are similar to (but not identical with) those previously described in consensus-type Zn fingers (derived from ADR1 and Xfin); the similarities suggest that even and odd ZFY domains bind DNA by a common mechanism. A model of the protein–DNA complex (designated the "jumping-linker" model) is presented and discussed in terms of the ZFY two-finger repeat. In this model *every other* linker is proposed to cross the minor groove by means of a putative finger/linker submotif HX_4HX_3–hydrophobic residue–X_3. Analogous use of a hydrophobic residue in a linker that spans the minor groove has recently been described in crystallographic and 3D NMR studies of homeodomain–DNA complexes. The proposed model of ZFY is supported in part by the hydroxyl radical footprint of the TFIIIA–DNA complex [Churchill, M. E. A., Tullius, T. D., & Klug, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5528–5532].

The Zn-finger motif defines a highly conserved class of eukaryotic nucleic acid binding proteins (Klug & Rhodes, 1987; Evans & Hollenberg, 1988). In this paper we describe 2D NMR[1] studies and structural modeling of an isolated Zn finger from the human male associated protein ZFY. This gene, originally identified from studies of sex reversal in man (de la Chapelle, 1972; Page et al., 1987), encodes a putative transcription factor proposed to participate in spermatogenesis (Palmer et al., 1989; Koopman et al., 1989). Genetic studies of intersex abnormalities of the newborn suggest that ZFY also plays an accessory role in the pathway of male sexual development in embryogenesis (Page et al., 1990).

Two-dimensional NMR studies have previously been conducted of single-finger peptides from *Saccharomyces* protein ADR1 and *Xenopus* protein Xfin (Parraga et al., 1988; Lee et al., 1989a,b). These fingers fold as compact globular minidomains in which the divalent metal is encaged. The N-terminal portion of the finger, containing the conserved cysteines, forms a β-sheet and β-turn (β-hairpin); the C-terminal

†Supported by the Lucille Markey Charitable Trust and by grants from the National Institutes of Health (HD 26465), the American Cancer Society, and the Whitaker Foundation to M.A.W. M.A.W. is supported in part by the Pfizer Scholars Program for New Faculty and a Junior Faculty Research Award from the American Cancer Society. T.H. is supported in part by NIH Grant GM 38221. M.K. is supported in part by a NATO postdoctoral fellowship.

*Address correspondence to this author at the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School.
‡Harvard Medical School.
§Permanent address: CNRS URA 1254, BIOP Polytechnique, 91128 Palaiseau, France.
‖Massachusetts General Hospital.

[1] Abbreviations: CD, circular dichroism; DQF-COSY, double-quantum-filtered correlated spectroscopy; DG, distance geometry; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; RMS, root mean square; SA, simulated annealing; TOCSY, total correlation spectroscopy.

0006-2960/91/0430-3371$02.50/0

portion, containing the conserved histidines, forms an α-helix ($\beta\beta\alpha$ motif). These features are in accord with earlier structural models (Berg, 1988; Gibson et al., 1988). The putative DNA-binding domain of ZFY (Page et al., 1987) is distinguished from the general class of Zn-finger proteins by the presence of a two-finger repeat (Figure 1). In this repeat the odd-numbered domains (fingers 1, 3, 5, 7, 9, 11, and 13) fit the general Zn-finger consensus sequence (Gibson et al., 1988), whereas the even-numbered domains (fingers 2, 4, 6, 8, 10, and 12) contain systematic differences. Circular dichroism studies suggest that peptides derived from odd- or even-numbered domains exhibit similar but not identical patterns of metal-dependent folding (Weiss et al., 1990). In this paper we will focus on the solution structure of a representive even-numbered domain.

Our results are presented in two parts. Part I describes the design of a 30-residue peptide (designated ZFY-6T)[2] as a model system and the sequential assignment of its 2D NMR spectrum. The observed pattern of nuclear Overhauser enhancements (NOEs) and *J*-coupling constants demonstrates that this variant finger retains the β-sheet/α-helix motif as previously described (Parraga et al., 1988; Lee et al., 1989a,b; Klevit et al., 1990). Long-range NOEs involving conserved framework residues define a hydrophobic core, and observation of slowly exchanging amide resonances indicates that this core is exceptionally stable. The three-dimensional structure of ZFY-6T is obtained in part II by analysis of NOE and *J*-coupling restraints using distance geometry and simulated annealing (DG/SA). The overall pattern of folding is similar to consensus-type Zn fingers; well-defined interactions are observed involving residues conserved among Zn fingers or particular to the subfamily of ZFY-related sequences. Our results suggest that odd- and even-numbered fingers share a common mechanism of DNA binding. A model of the ZFY protein–DNA complex (designated the "jumping-linker" model) is presented that retains the two-finger repeat as an essential feature. Aspects of this model are supported by a reinterpretation of the high-resolution hydroxyl radical footprinting of the TFIIIA–DNA complex (Churchill et al., 1990) and discussed in reference to analogous features of the homeodomain–DNA complex (Otting et al., 1990; Kissinger et al., 1990).

Materials and Methods

Peptide Synthesis and Characterization. ZFY-6P (sequence KPYQCQYCEYRSADSSNLKTHIKTKHSKEK) and ZFY-6T (sequence KTYQCQYCEYRSADSSNLKTHIKTKHSKEK) were synthesized by the solid-phase procedure (Barany & Merrifield, 1979; Stewart & Young, 1984) and purified following reduction (below) by reversed-phase HPLC as previously described (Weiss et al., 1990). Quantitative ninhydrin tests were used to monitor coupling efficiency, which averaged 99.1–99.3%. The following protecting groups were used on the *t*-BOC amino acids: chlorobenzoxy (Lys), carbobenzoxy (His), bromobenzoxy (Tyr), methylbenzyl (Cys), benzyl (Ser, Thr, Glu, Asp), and tosyl (Arg). Purity was evaluated by analytical HPLC, composition, and sequence analysis of preview (Tregear et al., 1977). Peptides were also prepared by using F-moc methodology as previously described (Weiss et al., 1990).

Reduction of Cystine to Cysteine. This was accomplished by reaction with 0.5 M dithiothreitol in 100 mM Tris-HCl (pH 7.7 at 20 °C) at 60 °C for 1–3 h (Frankel et al., 1987). The reduction status of cysteines was confirmed by reaction with iodoacetate followed by sequencing. The reduced peptide exhibited a different elution position in reversed-phase HPLC than an oxidized form containing an intramolecular disulfide bond, thus enabling the reduced and oxidized forms to be characterized individually (Weiss & Keutmann, 1990).

Aggregation State. This was determined by gel filtration chromatography (Sephadex G-50 Fine). The elution position expected for a monomeric peptide was calibrated in reference to a fragment of parathyroid hormone (residues 1–34). The oligomeric states of the ZFY-6P/Zn^{2+} and ZFY-6T/Zn^{2+} complexes were found to be monomeric at the concentration and conditions of study (see part I under Results).

NMR Sample Preparation. A total of 12 mg of the reduced peptide was dissolved in 0.7 mL of deoxygenated NMR buffer (see below). The concentration of peptide was ca. 5 mM. Following NMR data collection, the sample was lyophilized and stored in vacuo; under such conditions the ZFY-6T/Zn^{2+} complex did not undergo significant oxidation (<5%) for several months.

Buffers. Absorption and CD spectra were performed in 50 mM Tris-HCl (pH 7.5) containing successive aliquots of $CoCl_2$ or $ZnCl_2$. For pH titrations this buffer was mixed with aliquots of 0.1% acetic acid (containing the same concentration of $CoCl_2$ or $ZnCl_2$) to achieve intermediate pH conditions. For NMR measurements deuterated Tris-HCl was used. To delay oxidation of the peptide, buffers were purged with Ar or N_2 immediately prior to use.

NMR Methods. Spectra were recorded at 500 MHz at the Harvard Medical School NMR center. Preliminary spectra were obtained at the Columbia College of Physicians and Surgeons; Francis Bitter National Magnet Laboratory, MIT; NMR Facility of the University of Wisconsin, Madison; and Laboratory of Chemical Physics, National Institutes of Health. Two-dimensional experiments were obtained at 25 °C by the pure-phase method of States et al. (1982) and ordinarily acquired with 4K points in the t_2 dimension and 600 t_1 increments. Data were zero-filled to 4K in the t_1 dimension; exponential and shifted sine-bell window functions were applied in both dimensions prior to Fourier transformation.

J-coupling constants were determined from analysis of high-resolution DQF-COSY spectra; these were acquired with 8K points in the t_2 dimension. The processed spectrum, consisting of an 8K by 2K matrix, contained a digital resolution of 0.7 Hz/point in the ω_2 dimension. No line-width correction was applied to the apparent *J*-values in the antiphase multiplets; these should therefore be considered as upper limits to the actual coupling constants.

NMR-Derived Restraints. NOE and *J*-coupling (dihedral angle) restraints were used for molecular modeling. (i) NOEs (mixing times of 100 and 200 ms) were classified as strong (<2.7 Å), medium (2.8–<3.3 Å), or weak (3.4–<4.3 Å) on the basis of relative amplitudes of two-dimensional cross peaks; the intraresidue ortho–meta cross peak of tyrosine (2.5 Å) and the H_δ–H_ϵ cross peak of histidine (4.3 Å) were used as internal standards. Most NOEs were observed in both 200- and 100-ms mixing-time NOESY spectra. Weak NOEs observed at a mixing time of 200 ms, but not 100 ms, were generally regarded as spin diffusion effects and not used (in selected cases a few such NOEs were used when no spin diffusion

[2] Nomenclature: In this paper ZFY-6P denotes the 30-mer peptide containing proline at position 2, ZFY-6T denotes the corresponding peptide with threonine at position 2, and ZFY-6 denotes in generic terms the even-specific motif of domain 6. Odd-numbered linkers are defined as those that follow odd-numbered ZFY domains; even-numbered linkers are defined as those that follow even-numbered ZFY domains. Amino acids are designated by standard single-letter code.

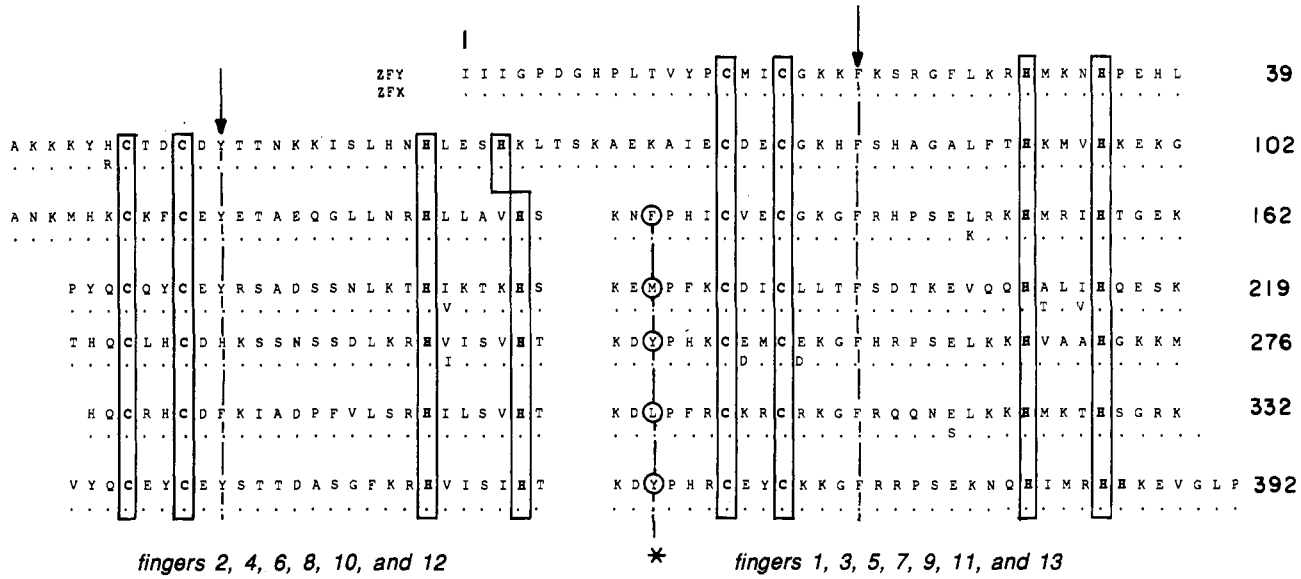

FIGURE 1: Sequence of the putative DNA-binding domain of the human ZFY and ZFX Zn-finger proteins. The two-finger repeat is shown in an alignment developed by Page et al. (1987). Cysteines and histidines presumed to be involved in Zn^{2+} coordination are boxed, conserved hydrophobic residues in even-numbered linkers are encircled (asterisk), and the alternating positions of central aromatic residues are indicated by arrows.

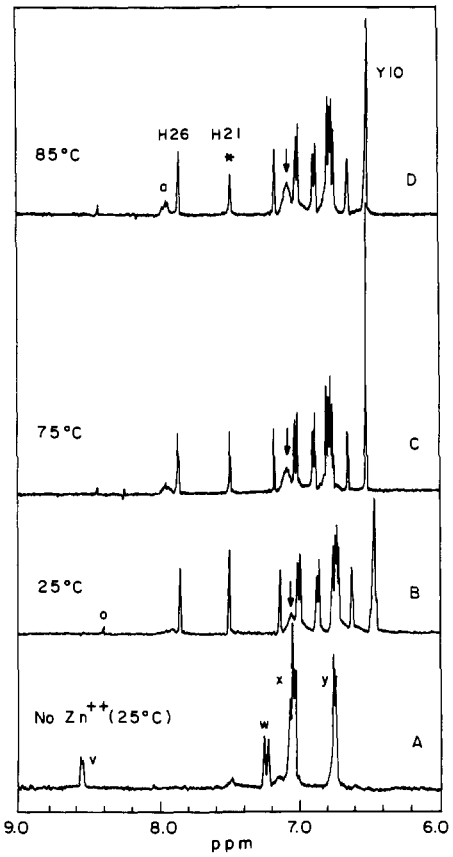

FIGURE 2: One-dimensional ^1H NMR spectra of the aromatic region of ZFY-6T in D_2O in the absence of divalent metal (panel A) and in the presence of equimolar Zn^{2+} (panels B–D). In panel A resonances v and w are assigned to the H_ϵ and H_δ resonances of the histidines; resonances x and y are assigned to the respective ortho and meta protons of the tyrosines in the unfolded peptide. A broad resonance (indicated by arrows in panels B–D) is tentatively assigned to the resonance of Y10 in an alternative configuration (see text). Resonance o is an impurity. Resonances labeled a in panel D are due to aggregated oxidized peptide (less than 5% of the material).

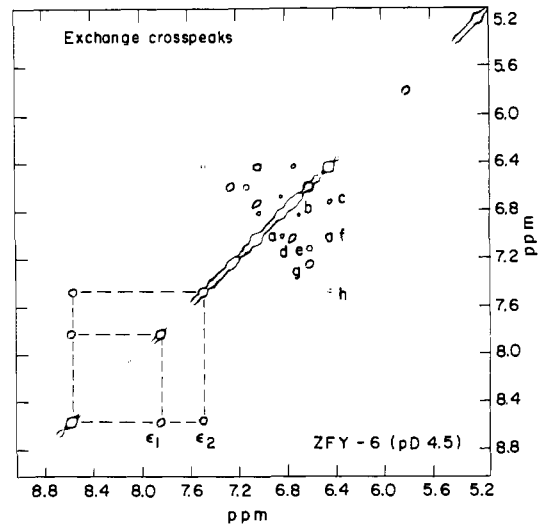

FIGURE 3: NOESY spectrum of the ZFY-6T/Zn^{2+} complex at pD 4.5 demonstrating exchange cross peaks between folded (metal bound) and unfolded (metal free) states (cross peaks ϵ_1 and ϵ_2 are assigned to the H26 and H21 H_δ protons, respectively; cross peaks c and f are assigned to the meta and ortho protons of Y10; cross peak g is assigned to the H_ϵ proton of H26; cross peak a is assigned to the ortho proton of Y7). Cross peaks b and d are assigned to the interresidue ortho–meta NOEs of Y7 and Y3, respectively, in the folded state. Cross peaks e and h are characteristic of the folded state and are assigned to NOEs between H_δ–H21 and H_δ–H26 and between Y10 (unresolved aromatics) and H_ϵ–H21, respectively. Experimental conditions: mixing time 200 ms and 25 °C. This spectrum was recorded at 400 MHz at the NMR Center of the Columbia College of Physicians and Surgeons (courtesy of Prof. D. J. Patel).

pathway was apparent in the initial DG/SA model). Distance-bound corrections were made for methyl groups, the degenerate ring protons of tyrosine, and methylene protons for which stereospecific assignments could not be obtained (Wuthrich, 1986). A correction of 3.4 Å was applied for NOEs involving the unresolved ortho and meta proton of Y10 (see part II under Results). (ii) Dihedral angle restraints were introduced on the basis of J-coupling constants in accord with the Karplus equation (Karplus, 1959) as previously described (Pardi et al., 1984). ϕ dihedral angles of residues with small $^3J_{HN\alpha}$ coupling constants (<5.5 Hz) were constrained between −90° and −40°; those with large $^3J_{HN\alpha}$ coupling constants (>8.5 Hz) were constrained between −160° and −80°. Restraints were not introduced on the basis of intermediate J

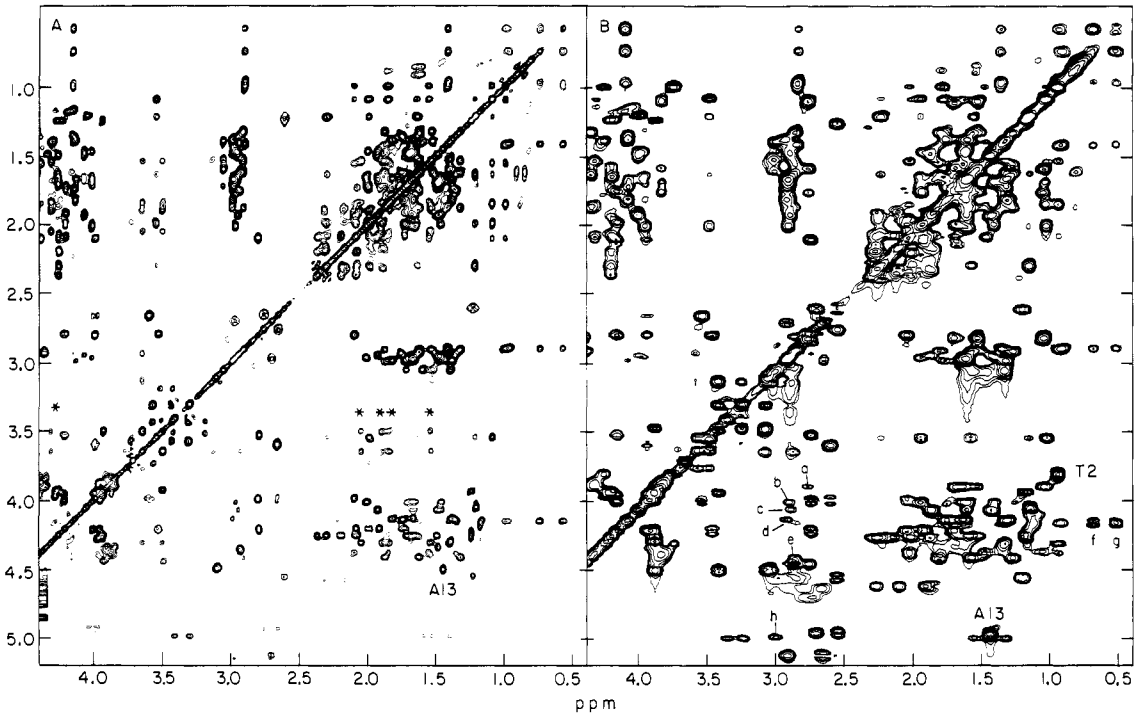

FIGURE 4: TOCSY spectra of ZFY-6P (panel A) and ZFY-6T (panel B) that exhibit similar dispersion of aliphatic spin systems. Resonances significantly affected by the P2T substitution are adjacent in the primary structure (residues 1 and 3) or adjacent in the tertiary structure (Ala 13; labeled in both panels). Representative connectivities of the proline spin system are indicated by an asterisk. In panel B representative connectivities of lysines or arginine are labeled as follows: (a) K1 H_α–H_ϵ; (b) K23 H_α–H_ϵ; (c) K19 H_α–H_ϵ; (d) K25 H_α–H_ϵ; (e) K30 H_α–H_ϵ; (f) K25 H_α–$H_{\beta 3}$; (g) K25 H_α–$H_{\beta 2}$; (h) R11 H_α–H_δ. Experimental conditions: 55 ms mixing time, 25 °C, and pH 6.2. Spectrum A was recorded at 600 MHz at the NIH Laboratory of Chemical Physics (courtesy of Dr. A. Bax) and spectrum B at 500 MHz at the Harvard Medical School NMR Center.

values. χ_1 dihedral angles, when obtained, were restrained about the preferred rotamer with a tolerance of ±30°. A similar restraint was introduced for the χ_2 dihedral angle of L18, for which stereospecific assignment of the methyl resonances was obtained (Results). A table of experimental restraints is provided as supplementary material (see paragraph at end of paper).

Distance Geometry/Simulated Annealing. Calculations were performed with a new set of DG programs, designated DG-II (Havel, 1991). This set is based on the EMBED algorithm (Crippen, 1981; Havel et al., 1983; Crippen & Havel, 1988) and incorporates recent improvements in sampling and convergence (Havel & Wuthrich, 1984; Havel, 1990). When applied to NOE and *J*-coupling information for ZFY-6T, DG-II provided an ensemble of 10 structures; two of the calculations did not converge and were not considered further. None of the remaining eight structures violated any NOE constraint by more than 0.26 Å (in seven of the structures there were no NOE violations greater than 0.16 Å).

RESULTS

(I) 2D NMR Studies of ZFY-6P and ZFY-6T

Design of the Model System. The odd-numbered domains in the ZFY-related gene family are similar to the general Zn-finger template (Gibson et al., 1988):

(N)hPAr-C--C-K-F----EL--H---H-A/G-
1 5 10 15 21 25

where h denotes a hydrophobic residue and Ar denotes an aromatic group (His, Phe, or Tyr). In contrast, the even-numbered domains follow the variant consensus:

(N)K-ArQC-Y/HC-F/Y-S/T----EL--H----HS/TK
1 5 10 15 21 26

The even consensus is distinguished by a conserved glutamine N-terminal to the first cysteine (Q4), the position of the central aromatic group (Y10 or F10, underlined above and indicated by arrows on Figure 1), and the spacing between histidine ligands (HX_4H). Each of these distinguishing features will be seen to play a well-defined role in the 2D NMR structure of an even-numbered domain (part II). ZFY-6P and a related analogue ZFY-6T were chosen as NMR model systems following systematic screening of single-finger peptides as appropriate systems for NMR study (Weiss et al., 1990). Their sequences derive from residues 162–191 (domain 6) of the putative DNA-binding domain of ZFY (Figure 1).

The C-terminal residue of ZFY-6P (corresponding to Met191 in the exon) is replaced by lysine; this modification is meant to enhance solubility and avoid heterogeneity due to methionine oxidation. ZFY-6T contains an additional modification: the second residue (corresponding to Pro163) is replaced by threonine. Among ZFY-related sequences Thr2 is more characteristic than Pro2 of the even consensus. In the course of screening analogues of ZFY-6 (Weiss & Keutmann, 1990), we observed that the Thr2 finger (designated ZFY-6T) is more stable than the parent peptide under acidic conditions (the pH midpoints for unfolding of the ZFY-6P/Co^{2+} and ZFY-6T/Co^{2+} complexes are 5.8 and 5.3, respectively). The ZFY-6P/Zn^{2+} and ZFY-6T/Zn^{2+} complexes are monomeric under the conditions of study, as indicated by gel filtration chromatography (Figure A of supplementary material).

Metal-Dependent Folding. Characteristic charge-transfer and *d–d* bands in the visible absorption spectra of the ZFY-6P/Co^{2+} and ZFY-6T/Co^{2+} complexes demonstrate tetrahedral metal coordination at pH 7 as previously described (Weiss et al., 1990). Circular dichroism spectra of these and related analogues exhibit Co^{2+}- and Zn^{2+}-dependent formation of α-helical structure (Weiss et al., 1990; Weiss & Keutmann, 1990). Corresponding changes are observed in the 1D ^1H

Table I: Chemical Shifts of the Assigned ^{1}H NMR Resonances of ZFY-6T at pH 6.5 and 25 °C

	chemical shifts at 25 °C					
residue	NH	$C_\alpha H$	$C_\beta H$			others
K1		3.83	1.68		1.63	$C_\gamma H_2$ 1.05; $C_\delta H_2$ 1.52; $C_\epsilon H_2$ 2.75
T2	8.14	4.25		3.75		$C_\gamma H_3$ 0.95
Y3	8.8	4.4	2.85		2.75	$C_{2,6}H$ 6.88; $C_{3,5}$ 6.72
Q4	8.46	4.55	1.92		1.85	$C_\gamma H_2$ 2.11, 2.28
C5	8.9	4.15	3.48		2.75	
Q6	8.82	4.18		1.58		$C_\gamma H_2$ 1.52, 1.70
Y7	9.65	4.5	2.57		1.2	$C_{2,6}H$ 6.84; $C_{3,5}$ 6.69
C8	7.9	4.93	3.35		3.23	
E9	8.46	4.18	2.02		1.88	$C_\gamma H_2$ 2.25, 2.13
Y10	9.08	3.93	2.73		2.05	$C_{2,6}H$ 6.42; $C_{3,5}$ 6.42
R11	6.91	4.92	1.58		1.32	$C_\gamma H_2$ 1.48, 1.48; $C_\delta H_2$ 3, 3
S12	8.39	4.45	3.88		3.43	
A13	8.65	4.9		1.45		
D14	8.21	4.88	2.7		2.5	
S15	8.65	3.25	3.04		3.4	
S16	8.33	4.13		3.95		
N17	8.17	4.42		3.05		
L18	7.49	4.32	2.03		1.48	$C_\gamma H$ 1.78; $C_\delta H_3$ 1.02, 0.93
K19	7.96	3.95		1.95		$C_\gamma H_2$ 1.4; $C_\delta H_2$ 1.6; $C_\epsilon H_2$ 2.92
T20	8.10	3.86		4.2		$C_\gamma H_3$ 1.16
H21	8.26	3.92	3.55		2.6	C_2H 7.47; C_4H 7.11
I22	8.65	3.48		1.92		$C_\gamma H_3$ 1.03; $C_\gamma H_2$ 1.58, 1.25; $C_\delta H_3$ 1.17
K23	7.72	4		1.85		$C_\gamma H_2$ 1.42; $C_\delta H_2$ 1.6; $C_\epsilon H_2$ 2.88
T24	7.87	3.96		3.96		$C_\gamma H_3$ 1.15
K25	8.17	4.06	0.48		0.67	$C_\gamma H_2$ 0.67; $C_\delta H_2$ 0.93; $C_\epsilon H_2$ 2.85
H26	7.37	5.07	2.92		2.65	C_2H 7.82; C_4H 6.59
S27	7.64	4.25		3.88		
K28	8.03	4.33	1.8		1.57	$C_\gamma H_2$ 1.26; $C_\delta H_2$ 1.48; $C_\epsilon H_2$ 2.8
E29	8.06	4.15	1.85		2	$C_\gamma H_2$ 2.15, 2.25
K30	7.8	4.08	1.75		1.65	$C_\gamma H_2$ 1.3; $C_\delta H_2$ 1.6; $C_\epsilon H_2$ 2.92

NMR spectrum of the peptide upon addition of these divalent cations. In the absence of divalent metal the NMR spectrum of ZFY-6P or ZFY-6T resembles that of a random coil (panel A of Figure 2). In contrast, the 1D ^{1}H NMR spectrum of the Zn^{2+}–peptide complex exhibits marked dispersion of chemical shifts, reflecting a folded globular structure. This is illustrated for the aromatic resonances in panel B of Figure 2. At pH >5.5 a single spin system is observed in the fingerprint region for each residue (Figure 5; see below). This is consistent with the existence of a single conformation of the metal-folded peptide (an apparent minor conformer of the Y10 side chain is discussed below). The structure remains folded up to 85 °C (panels B–D). Similar metal-dependent ^{1}H NMR changes are observed upon addition of Cd^{2+} (data not shown).

pH-Dependent Unfolding. Coordinate loss of metal-binding and secondary structure has previously been described in the ZFY-6P/Co^{2+} complex between pH 5 and 6 (Weiss et al., 1990; Weiss & Keutmann, 1990); similar pH behavior has been observed in other systems (Parraga et al., 1990). In CD studies the ZFY-6P/Zn^{2+} complex exhibits similar loss of metal-dependent structure between pH 4 and 5; the shift in pH midpoint reflects the greater stability of the Zn^{2+} complex relative to Co^{2+}. pH-dependent folding and unfolding may also be monitored by ^{1}H NMR. Spectra obtained at the midpoint of the pH-unfolding curve demonstrate slow exchange between folded and unfolded species, reflecting a two-state process (Figure 3). In the NOESY spectrum exchange cross peaks are observed between liganded (H26 and H21 at 7.44 and 7.84 ppm, respectively) and unliganded (8.6 ppm) histidine $C_\epsilon H$ resonances (labeled ϵ_1 and ϵ_2 in Figure 3).

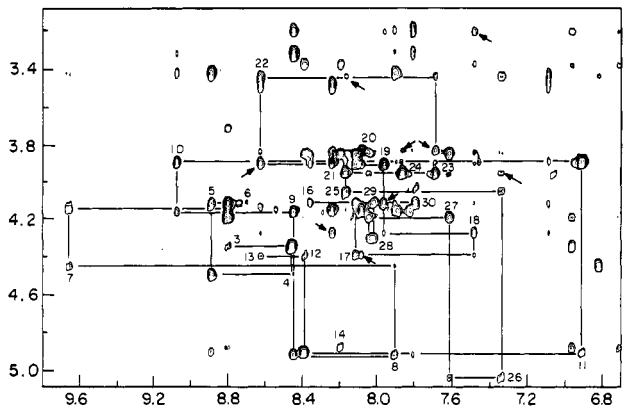

FIGURE 5: $d_{\alpha N}(i,i+1)$ connectivities of ZFY-6T in the NOESY spectrum recorded in 90% H_2O (conditions: 250-ms mixing time, pH 6.2, and 25 °C). Cross peaks are labeled by residue number. Helix-related $d_{\alpha N}(i,i+3)$ connectivities in the C-terminal domain are indicated by arrows (see Figure 7).

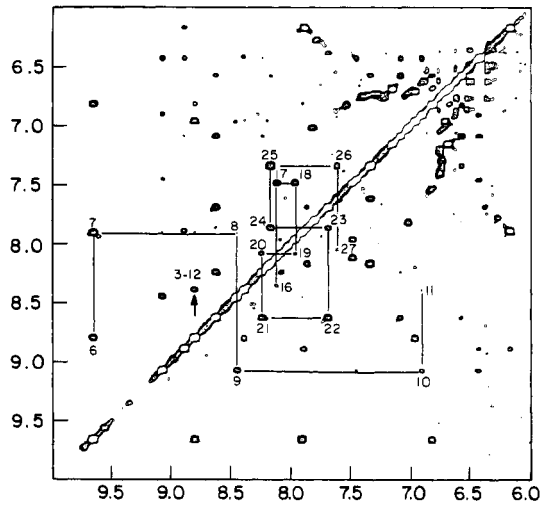

FIGURE 6: $d_{NN}(i,i+1)$ connectivities of ZFY-6T in the NOESY spectrum recorded in 90% H_2O (conditions: 250-ms mixing time, pH 6.2, and 25 °C). Cross peaks are labeled by residue number. The arrow indicates an interstrand NOE between residues 3 and 12 in the β-sheet.

As expected, NOEs characteristic of the folded state—as between, for example, the $C_\delta H$ protons of the liganded histidines (cross peak e) or between the ring protons of Y10 (unresolved at 6.42 ppm; see below) and the $C_\epsilon H$ proton of H21 (cross peak h in Figure 3)—are retained under these conditions. Two-state behavior has also been described in the pH-dependent unfolding of ADR1a (Parraga et al., 1990).

NMR Spin-System Identification. ZFY-6P and ZFY-6T each contain 30 residues. Their constituent spin systems may be classified as (i) methyl containing (A, I, L, and T), (ii) long methylene chains (K, R, Q, and E), (iii) aromatic residues (Y, H), and (iv) remaining AMX spin systems (C, S, N, and D). In this section these spin systems in ZFY-6T are described in turn; the spectrum of ZFY-6P is similar and will be described below.

(*i*) *Methyl-Containing Spin Systems.* ZFY-6T contains three threonines (positions 2, 20, and 24) and unique alanine (position 13), leucine (18), and isoleucine (22) residues. Their characteristic spin systems are readily identified in DQF-COSY and TOCSY spectra.

(*ii*) *Long Side Chains.* Eleven out of the 30 residues of ZFY-6T have long side chains. The six lysines (positions 1, 19, 23, 25, 28, and 30) and unique arginine (position 11) are readily identified in the TOCSY spectrum (Figure 4); in each

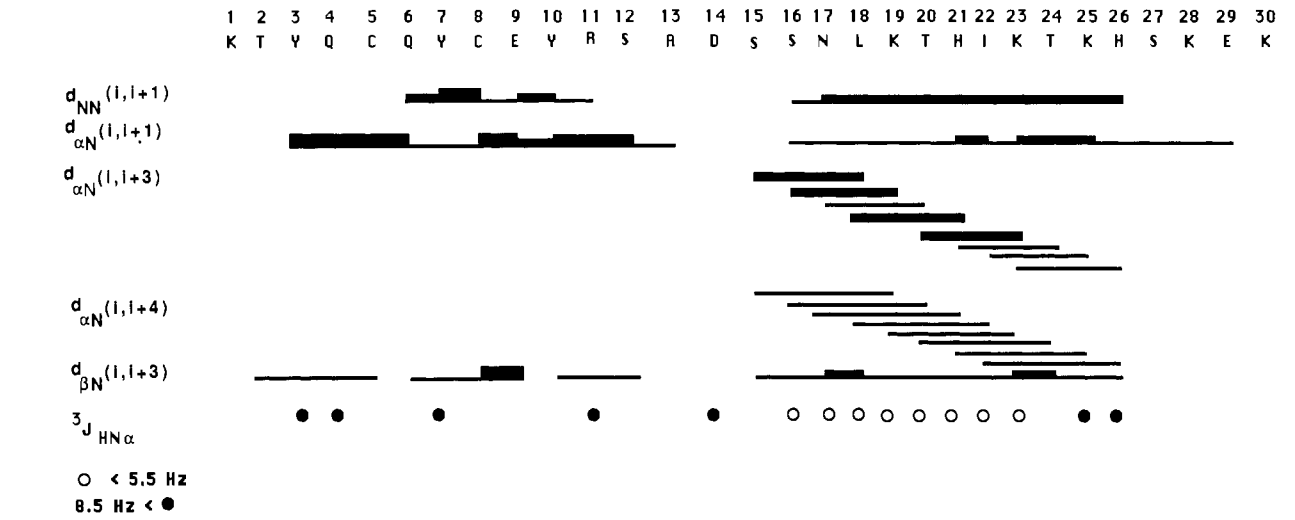

FIGURE 7: Schematic representation of the sequential connectivities and backbone *J*-coupling constants of ZFY-6T. Symbols and format are as described (Wuthrich, 1986).

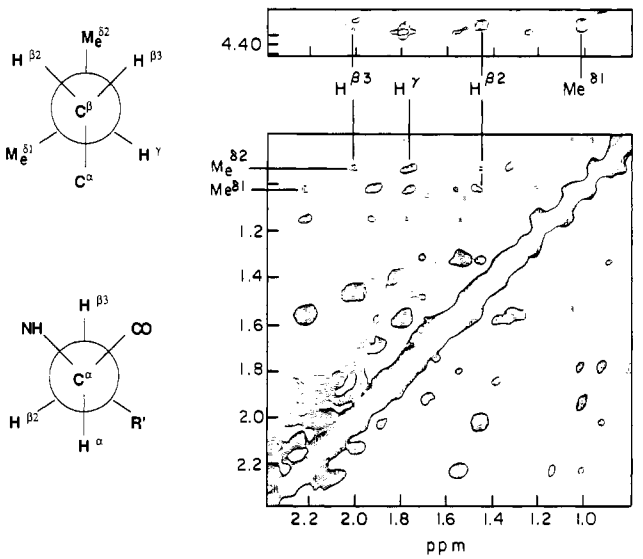

FIGURE 8: Stereospecific assignment of L18 obtained by an extension of the method of Wagner et al. (1987) for β-protons. Following stereospecific assignment of the β protons (summarized in a schematic form in the lower left panel), additional intraresidue NOEs [i.e., strong NOE cross peaks from H_α to H_γ and $C_{\delta 1}H_3$ (labeled $Me^{\delta 1}$), from $C_{\delta 2}H_3$ (labeled $Me^{\delta 2}$ in the figure) to both β protons, and from $H_{\beta 2}$ to both methyl resonances] make possible the stereospecific assignment of the L18 methyl groups (summarized in a schematic form in the upper left panel; see also text). Experimental conditions: 100-ms mixing time, pH 6.2, and 25 °C.

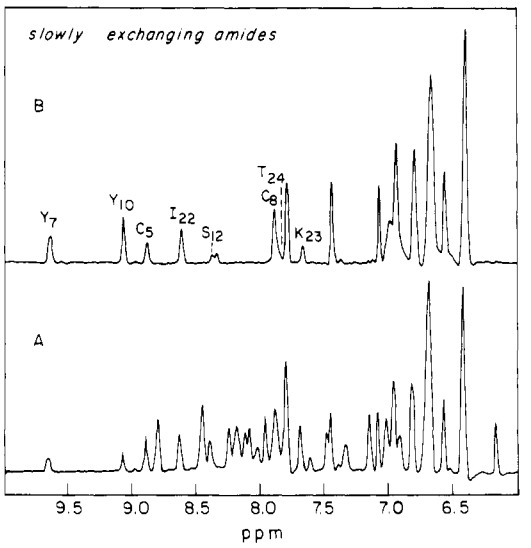

FIGURE 9: Slowly exchanging amide resonances of ZFY-6T at pD 6.0 (direct meter reading) demonstrating the stability of the Zn finger as a globular minidomain. Spectrum A was recorded 15 min after dissolution of a protonated sample in D_2O. Assignments are indicated (see Table I). Progressive exchange is observed over 24 h under these conditions. Spectrum B is a reference spectrum of the same sample in 90% H_2O/10% D_2O.

case connectivities are observed from H_α to each successive proton, including δ protons of arginine (cross peak h in panel B of Figure 4) and ϵ protons of lysine (cross peaks a-e). Four additional spin systems are observed due to glutamine or glutamic acid, which cannot be further distinguished at this stage of the analysis.

(*iii*) *Aromatic Residues.* The aromatic region of ZFY-6T is well resolved. The ortho and meta resonances of two of the three tyrosines are classified on the basis of their connectivities in the DQF-COSY and NOESY spectra. The H_ϵ protons of the two histidine in the folded peptide are assigned from their exchange cross peak with the same protons of the unfolded peptide at pH 4.5 (see above). The H_δ protons are then assigned from their strong TOCSY and weak intraresidue NOE connectivity to the H_ϵ. An additional composite resonance (6.4 ppm) is observed in D_2O, which upon integration contains approximately four proton resonances; this is assigned to the unresolved ortho and meta protons of the third tyrosine. In each case the histidine H_δ and tyrosine ortho resonances exhibit strong NOEs to the β protons of a single AMX spin system, enabling assignment of the residue side chains to be completed.

(*iv*) *Other AMX Spin Systems.* Eight remaining AMX spin systems (corresponding to four Ser, two Cys, one Asp, and one Asn; see Table I) are identified but not further distinguished at this stage.

Sequential Assignment. The NH–C_αH fingerprint region is well resolved in the NOESY spectrum of the Zn^{2+} complex (Figure 5). $C_\alpha H_i$ to NH_{i+1} connectivities are observed from residues 2 to 14 and from residues 16 to 30. In addition, amide–amide connectivities are observed from residues 6 to 11 and 16 to 27, as outlined in Figure 6. This network of connectivities provides the sequential assignment of residues 3 to 14 and 16 to 30. At pH 6–6.5 the NH resonances of T2 and S15 are exchanging too rapidly with water to give rise to observable resonances: these resonances are assigned on the basis of a DQF-COSY spectrum recorded at pH 5.5. The N-terminal NH_2 resonance of K1 is not observed under these

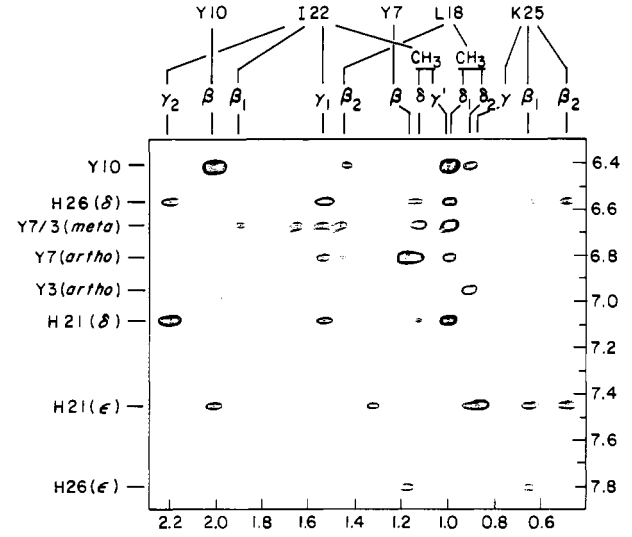

FIGURE 10: NOESY cross peaks between the aromatic protons of the histidines and tyrosines (vertical axis; ω_1) and methyl protons (horizontal axis; ω_2) demonstrating tertiary interactions in the hydrophobic core of the domain. Assignments are indicated (see Table I). The mixing time was 200 ms. Stereospecific NOEs indicate the absence of significant spin diffusion under these conditions.

conditions. A summary of sequential connectivities is shown in schematic form in Figure 7; ^{1}H NMR assignments of ZFY-6T are given in Table I. The assignment of two AMX spin systems as ligands (C5 and C8) is verified by observation of heteronuclear ^{113}Cd coupling (data not shown).

Stereospecific Assignments. Stereospecific assignments have been obtained for the β protons of 10 residues on the basis of (i) the value of the $C_\alpha H$–$C_\beta H$ coupling constants, (ii) the relative intensities of the intraresidue $C_\alpha H$–$C_\beta H$ NOEs, and (iii) the relative intensities of the NH–$C_\beta H$ NOEs. Stereospecific assignments have been used for those residues for which all criteria are compatible with one of the three preferred side-chain rotamers (Wagner et al., 1987). No assignments are made for the residues for which one of the coupling constants is in the range of 6–9 Hz or for which intraresidue NOE intensities are incompatible with the apparent coupling constants.

Once the stereospecific assignments of the β protons have been obtained, it is possible, in principle, to extend the previous procedure to the chemical groups attached to the C_γ carbon. This has been done in the case of Leu18 (Figure 8). The stereospecific assignment of the two δ-CH_3 resonances is based on their respective intraresidue NOE and coupling patterns: (i) strong NOEs from δ1-CH_3 to $C_\alpha H$ and $C_{\beta 2} H$ protons and absent NOEs from δ1-CH_3 to $C_{\beta 3} H$, (ii) strong NOEs from δ2-CH_3 to $C_{\beta 2} H$ and $C_{\beta 3} H$ protons and absent NOEs from δ2-CH_3 to $C_\alpha H$, and (iii) observation of a large coupling constant (>10 Hz) between $C_\gamma H$ and $C_{\beta 2} H$. (Due to resonance overlap the coupling constant between $C_\gamma H$ and $C_{\beta 3} H$ could not be obtained.) Rotamer relationships underlying this approach are depicted in schematic form in Figure 8.

Secondary Structure. The analysis of sequential and intermediate NOE patterns and of the NH–$C_\alpha H$ J-coupling constants (Figure 7) enables two regions in the structure of the finger to be distinguished: an N-terminal β-sheet and β-turn (β-hairpin) and a C-terminal α-helix ($\beta\beta\alpha$ motif). These elements of secondary structure are discussed in turn.

(i) N-Terminal β-Sheet. The N-terminal region, from residue 2 to residue 13, is characterized by two subdomains with strong $d_{\alpha N}$ NOE connectivities (2 to 6 and 8 to 12) and by strong $^3J_{HN\alpha}$ coupling constants in residues 3, 4, 7, and 11 (summarized in Figure 7). These features suggest the presence of two strands of extended chains. Interstrand NOEs observed between residues 2, 3, and 4 and residues 13, 12, and 11 (respectively) in turn indicate formation of a β-hairpin; the position of the turn is localized by d_{NN} NOE connectivities between residues 6 and 8. A representative interstrand NOE (between the HN resonances of Y3 and S12) is indicated by an arrow in Figure 6.

(ii) C-Terminal α-Helix. The C-terminal domain between residues 15 and 26 exhibits an NOE pattern [d_{NN}, $d_{\alpha N}(i,i+3)$, and $d_{\alpha N}(i,i+4)$] (Figure 7) characteristic of an α-helix. Additional evidence for α-helix formation is provided by small $^3J_{HN\alpha}$ coupling constants in residues 16 to 23 (<6 Hz). However, anomalously large $^3J_{HN\alpha}$ coupling constants are observed in residues 25 and 26, indicating a distortion of helical geometry at the C-terminal end of the α-helix. The structure of this distortion is defined by molecular modeling in part II (below).

Slowly Exchanging Amides. Several amide resonances are observed in freshly prepared D_2O solution (pD 6.0) and exchange with solvent over 10–24 h (Figure 9). These include the amide resonances of residues involved in stable secondary structure: C5, Y7, C8, Y10, and S12 in the β-sheet and I22, K23, and T24 in the α-helix. Each of the slowly exchanging protons is likely to be involved in a hydrogen bond; their exchange times presumably also reflect the stability of the hydrophobic core (below), as suggested in other systems by comparative NMR studies of engineered proteins (Jandu et al., 1990).

Tertiary Structure. Close packing between the N-terminal β-hairpin and C-terminal α-helix is demonstrated by the presence of several NOE connectivities between the aromatic protons of Y3, Y7, and Y10 and the side chains of L18, I22, and L18, respectively (Figure 10). These interactions are stereospecific; i.e., the two L18 methyl resonances exhibit distinct NOE patterns. NOEs involving the ortho and meta protons of Y10 cannot be distinguished, as these resonances (6.42 ppm; labeled in panels B–D of Figure 2) are not resolved. Accordingly, further definition of its interactions will require selective isotopic labeling, as discussed below The extensive set of observed β-hairpin/α-helix interactions defines a hydrophobic core which is similar to that observed in larger globular proteins; the three-dimensional structure of this core is defined in part II (below).

Zinc Coordination. Tetrahedral coordination to the two S_γ atoms of the cysteines and to the two histidines is supported by previous studies of Zn-finger domains (Lee et al., 1989a,b; Klevit et al., 1990). Both N_ϵ and N_δ coordination of the histidine are observed in the metal-binding sites of proteins; accordingly, in a variant finger motif one coordination scheme cannot be a priori excluded in favor of another. However, only N_ϵ coordination of both histidines is compatible with the observed NOE pattern: strong NOEs between the H_δ protons of the two histidines and between the H_ϵ proton of His26 and the β protons of Cys8 (Figure B of supplementary material). Similar conclusions have been obtained in 2D NMR studies of the ADR1 and Xfin Zn fingers (Lee et al., 1989a,b; Klevit et al., 1990).

Alternate Configuration of Y10. The aromatic spectra of ZFY-6P and ZFY-6T contain a set of broad resonances (arrow in Figure 2) whose integrated intensity at 25 °C is <10% of that expected for tyrosine (four protons). With increasing temperature, these broad resonances increase in relative intensity, apparently at the expense of the composite Y10 ring resonances at 6.42 ppm. These results suggest that the Y10 ring may occupy a second minor configuration. However, no NOEs are observable from the broad spin system, and so it

cannot be characterized further at this time. Accordingly, only the predominant conformer will be considered in part II. [As a control, we note that this NMR feature is identical in peptides synthesized by *t*-BOC or F-moc chemistries; since these chemistries differ in side-chain protecting groups, the additional resonances cannot be attributed to incomplete side-chain deprotection. We further note that this minor feature is absent in spectra of ZFY-6 analogues containing substitutions at position 10 (Weiss & Keutmann, 1990; unpublished results).]

Comparison of ZFY-6P and ZFY-6T. The above 2D NMR analysis was performed on ZFY-6T, a more stable analogue containing the substitution P2T. The 2D NMR spectrum of ZFY-6P is essentially the same, with the following exceptions: (i) A unique proline spin system is observed (indicated by asterisks in panel A of Figure 4). It adopts a unique configuration; no minor conformation is observed in slow exchange in the spectrum of ZFY-6P, suggesting by its absence the existence of local structure stabilizing one isomer (trans or cis) of the peptide bond. Isomer-specific NOEs to the preceding residue (Wuthrich, 1986; i.e., H_α and H_N of K1) are not readily observed, presumably due to disorder and rapid solvent exchange of the N-terminal NH_2 protons. (ii) The spin system of Thr2 in ZFY-6T is absent as expected (Figure 4). (iii) A nonlocal perturbation in chemical shift is observed in the resonances of Ala13 (i.e., across the β-sheet; labeled in Figure 4). (iv) An interstrand NOE is observed between the H_α of Ala13 and the H_α of Pro2 (data not shown), indicating that residue 2 participates in the β-sheet in both peptides.

(II) Three-Dimensional Structure of ZFY-6T

Experimental Restraints. A total of 135 NOE restraints were obtained; of these 63 were sequential (Figure 7), 33 were short range (between residues 2–4 positions apart in the sequence), and 33 were long range (>4 residues apart in the sequence). In addition, 8 intraresidue NOEs were used in selected cases to constrain side-chain configurations. Coupling constants were used to obtain eight ϕ, ten χ_1, and one χ_2 restraints. A table of the restraints is provided as supplementary material. The precision in the present structure calculation is limited by ambiguity among NOEs involving the aromatic ring of Y10, as the four ring protons have identical chemical shifts.

Zn^{2+}-Related Restraints. For DG/SA calculations the Zn^{2+}–S_γ distances were constrained to be 2.3 ± 0.1 Å and the Zn^{2+}–N_ϵ distances to 2.0 ± 0.1 Å; the zinc was constrained to lie in the plane of the imidazole rings. The bond angles of the Zn^{2+}-coordinating tetrahedron were constrained to be 109 ± 10°.

Distance Geometry/Simulated Annealing. Ten DG/SA structure calculations were initiated. Of these, two did not converge and were discarded. The remaining eight structures exhibited an RMS deviation for backbone atoms (residues 2–27) of 0.86 Å; the RMS deviation of all atoms (residues 2–27) was 1.93 Å. No experimental restraints were observed or provided for residues 1 and 28–30; these residues appear to be disordered in solution and will not be considered further. The ensemble of the eight DG/SA structures obtained is shown in green in panel A of Figure 11. The backbone configuration is well-defined and reveals an N-terminal β-hairpin and a C-terminal α-helix folded as a compact globular minidomain. Similar features have previously been described in isolated Zn fingers from Xfin-31 (Lee et al., 1989a,b) and ADR1 (Klevit et al., 1990).

Comparison with the Xfin-31 Zn Finger. Comparison of the ZFY-6T ensemble and a representative Xfin-31 structure (Lee et al., 1989a,b; coordinates obtained from D. Case and P. E. Wright, personal communication) is also shown in panel A of Figure 11 (ZFY-6T in green, Xfin-31 in red). The secondary structure and tertiary orientation of these two fingers are remarkably similar. Two prominent differences are observed. First, the ZFY-6T structures exhibit a distortion of α-helical geometry in the HX_4H metal-binding site; this is apparently required to maintain proper tetrahedral coordination of Zn^{2+}. Second, a difference is observed in the relative length of the β-turn; this does not appear to affect the linker/β-sheet junction. These differences notwithstanding, features proposed to be related to DNA recognition are similar in the two structures; these include the orientation of the fingertip and N-terminal portion of the α-helix (which may contain DNA-contact residues (Berg, 1988, 1990) and the spatial separation of the N- and C-termini (which defines possible finger–finger linkage arrangements in the DNA complex; see Discussion). Because the sequence of Xfin-31 is similar to that of odd-numbered ZFY domains, these results suggest that—whatever the structural role of the ZFY two-finger repeat—both odd- and even-numbered domains are likely to bind DNA by a common local mechanism. Possible models of the ZFY–DNA complex are presented below under Discussion.

Stabilization of the N-Terminal β-Hairpin. ZFY domains 6, 8, 10, and 12 contain glutamine just N-terminal to the first cysteine (Q4; see Figure 1). This residue is conserved among the chicken and mammalian ZFY-related sequences (DiLella et al., 1990), with the exception of a corresponding glutamic acid in domain 6 in the two Zfy alleles in the mouse (*M. dom.*). The two amide protons of Q4 exhibit inequivalent chemical shifts, are not attenuated following 1–2 s of solvent presaturation, and exhibit NOEs to the C5 amide and E9 H_α. These features suggest that the side chain participates in well-defined interactions. In the DG/SA structures Q4 is positioned to contribute hydrogen bonds to the carbonyl groups of C5 and E9, thus stabilizing the β-sheet and β-turn. Distortion of β-sheet geometry is not observed in ZFY-6T, unlike the solution structure of ADR1a (Klevit et al., 1990).

Structure of the Hydrophobic Core. The relationship between the N-terminal β-hairpin and C-terminal α-helix is defined by a well-ordered hydrophobic core. This is composed of hydrophobic and aromatic residues L18, Y3, Y7, Y10, and I22 and the methylene chain of K25; these residues pack against the invariant tetrahedron of ligands provided by C5, C8, H21, and H26. The core residues Y10, L18, and H21 are shown in panel B of Figure 11; their local environments are described in turn. The present description is qualitative; a more detailed analysis of packing interactions will require further refinement of the 2D NMR structure.

(*i*) *Leucine 18* is a conserved feature of classical Zn fingers and is observed to pack against Y10 and I22. L18 is occasionally replaced by A, V, or F in variant sequences. The function and folding of these variants have, to date, not been examined in this or other Zn-finger systems.

(*ii*) *Isoleucine 22.* A branched-chain aliphatic residue (I, L, or V) is conserved at this position in the HX_4H metal-binding motif among even ZFY-related domains. Interactions are observed involving L18 (above), Y7, H21, and H26. A greater variability at position 22 is observed among non-ZFY-related HX_4H fingers; the most common residues are Q and R. Multiple substitutions are also observed among odd-numbered ZFY domains and classical HX_3H fingers in general; Xfin-31 has Q at this position (Lee et al., 1989a,b).

(*iii*) *Lysine 25* adopts an unusually well-defined configuration. Its methylene chain packs near the two conserved

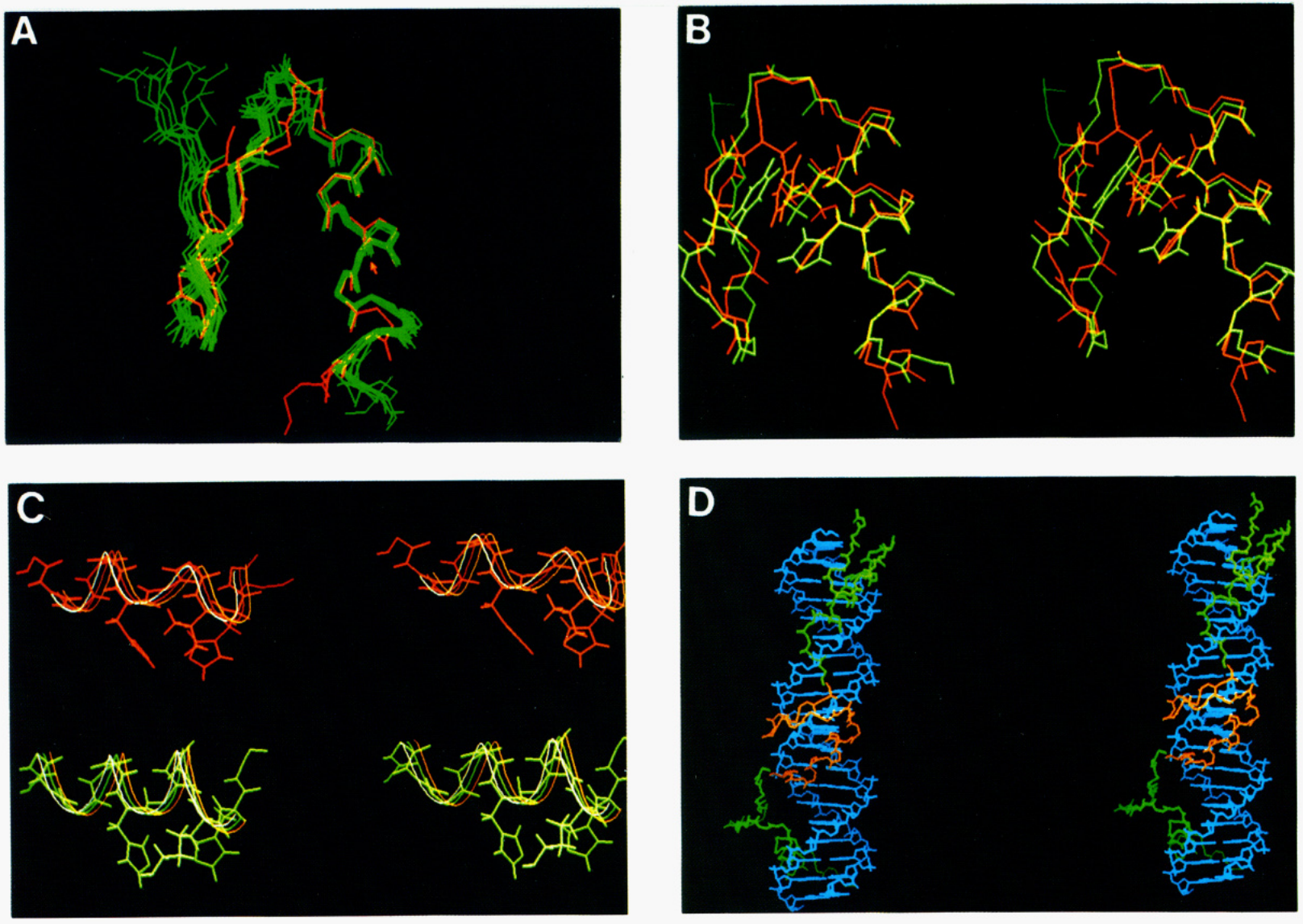

FIGURE 11: (A) Ensemble of eight DG structures of ZFY-6T (residues 2–27) showing the backbone configuration (green). For comparison, the average backbone configuration of Xfin-31 (Lee et al., 1989a,b) is shown in red (Brookhaven Data Bank). Residues 1 and 28–30 of ZFY-6T are omitted; no NMR-derived restraints were obtained for these residues, which are presumed to be disordered in solution. (B) Alternate aromatic residues in the hydrophobic core (Y10 in ZFY-6 and F12 in Xfin-31, using the ZFY-6 numbering scheme) exhibiting similar internal packing interactions in the two structures. In each case close contacts are observed with L18 and H21. The alternate sequence position of the central aromatic residue is a characteristic of the ZFY two-finger repeat (indicated by arrows in Figure 1). (C) C-Terminal α-helices containing the HX_4H (ZFY-6T) or HX_3H (Xfin-31) histidines involved in metal coordination. The distorted ZFY-6T helix is shown in the lower panel and the Xfin-31 helix in the upper panel. (D) Model of a ZFY–DNA complex exhibiting a two-finger structural repeat. A linker seven residues in length (i.e., HX_4H–X_7–$CX_{2-4}C$) is sufficient to span the minor groove in a direction approximately parallel to the axis of the double helix (see Figure 12).

histidine rings (panel C of Figure 11); the charged ϵ-amino group is exposed to solvent. The β-resonances (Figure 4B; α–β_3 and α–β_2 cross peaks are labeled f and g, respectively) exhibit a dramatic upfield shift (0.4 ppm; estimated secondary shift 2 ppm). This may arise from the superposition of ring-current effects from both imidazole rings (Johnson & Bovey, 1958). Xfin-31 has a valine at this position, which is also shifted upfield (Lee et al., 1989a,b). Among even ZFY-related domains, position 25 usually contains hydrophobic residues: valine in domains 4, 8, and 10 and isoleucine in domain 12 (the murine Zfy sequences contain alanine in domain 10). Interestingly, lysine is observed only in domain 6 and is conserved among avian and mammalian sequences. Among unrelated HX_4H domains, the above residues and phenylalanine are commonly observed; serine, threonine, and alanine are also seen. The functional implications of these sequence variations are not presently understood in ZFY or other Zn-finger systems.

(*iv*) *Aromatic Environments.* ZFY-6T contains three tyrosines (residues 3, 7, and 10), which are conserved as aromatic groups (Y, F, or H) among even-numbered ZFY domains. We have previously proposed that histidine substitutions at these positions destabilize the C-terminal α-helix and by so doing indirectly influence DNA sequence specificity (Weiss et al., 1990). Among Zn fingers generally an aromatic is observed at position 3, a hydrophobic or long-chain residue at position 7, and aromatic (F > Y) at alternative position 12; the latter pattern is exhibited by the odd-numbered ZFY domains (Figure 1).

Y3, Y7, and Y10 are observed to occupy distinct local environments. Y3 is located in the N-terminal strand of the β-sheet; its hydroxyl group may hydrogen-bond with the side chain of S15. Y7 is located between the two invariant cysteines in the loop of the β-hairpin and packs against I22 and H26. Y10 is located near the C-terminus of the β-sheet and packs against L18 and H21. Its hydroxyl group is positioned to hydrogen-bond to the side chains of either S12 or N17. Its interactions are essentially equivalent with those of the central phenylalanine in the 2D NMR structure of Xfin-31 [F12 in the present numbering scheme; residue 10 in the numbering scheme of Lee et al. (1989a,b)], as shown in panel B of Figure 11. Equivalence of these alternate aromatic positions was predicted by Berg (1988) and suggested by previous unfolding studies of an "aromatic swap" analogue (Weiss & Keutmann, 1990). The precision with which the configuration of Y10 is determined in the present 2D NMR structure is limited by ambiguities in the intepretation of its NOEs; as noted in part I, the ortho and meta ring resonances are unresolved. Further analysis will require selective isotopic labeling of this residue (Kochoyan et al., 1991).

Discussion

The ZFY gene family—encoding a subgroup of Zn-finger proteins distinguishable by a two-finger repeat—provides a model system for studying sequence-dependent variations in finger architecture and their implications for DNA recognition. Odd-numbered domains and linkers resemble the general Zn-finger consensus, whereas even-numbered fingers and linkers show systematic differences (Figure 1). Because these differences have been strictly conserved among vertebrates (DiLella et al., 1990), they are likely to reflect structural aspects of the putative protein–DNA complex. We have recently shown that odd and even ZFY domains exhibit similar but not identical patterns of metal-dependent folding, as monitored at low resolution by circular dichroism (Weiss et al., 1990). In this paper we provide further characterization of these patterns by 2D NMR analysis and computer-based modeling of a representative even-numbered domain (ZFY-6T). Similar 2D NMR studies have previously been described of consensus-type Zn fingers from other systems (Parraga et al., 1988; Lee et al., 1989a,b; Neuhaus et al., 1990). These studies are extended in the present work to examine a finger that shares some but not all of the features of the earlier systems.

Our results are discussed in two parts. General features of the structure of ZFY-6T are discussed in part I and compared to previously characterized classical Zn fingers (Lee et al., 1989a,b; Klevit et al., 1990). Implications for protein–DNA recognition by Zn-finger proteins are discussed in part II. A model of the ZFY–DNA complex (designated the "jumping-linker" model) is presented which incorporates a two-finger repeat. Remarkably, aspects of this model are supported by the recently described hydroxyl radical footprint of TFIIIA (Churchill et al., 1990).

(I) The ZFY Two-Finger Repeat: Structure of an Even-Numbered Domain

Each of the classical Zn fingers studied to date folds as a compact globular minidomain in which the divalent metal is encaged. The N-terminal portion of the finger, containing the conserved cysteines, forms a β-sheet and β-turn (hairpin); the C-terminal portion, containing the conserved histidines, forms an α-helix. The sequences of these fingers resemble those of the odd-numbered domains in the ZFY two-finger repeat (Page et al., 1987). Interestingly, a similar organization of secondary structure is observed in the solution structure of ZFY-6T ($\beta\beta\alpha$ motif; see panel A of Figure 11). Their higher order organization is stabilized by a hydrophobic core defined by the Zn^{2+} ligands and a conserved pattern of hydrophobic and aromatic residues; the latter include Y3, Y7, Y10, and L18 (using the ZFY-6 peptide numbering scheme). These features are in accord with earlier structural models (Berg, 1988; Gibson et al., 1988). We note in passing that the similarity of the solution structures of ZFY-6T and Xfin-31 (obtained independently by using different DG/SA protocols) provides an illustration that present NMR-based structure-determination procedures are well-defined.

Alternation of Aromatic Sites in the ZFY Two-Finger Repeat. The solution structure of ZFY-6T exhibits a number of features not observed in previous Zn-finger structures. The placement of the central aromatic residue (Y10; two residues C-terminal to the second cysteine) differs from that of a related aromatic residue in the general Zn-finger consensus (four residues C-terminal to the second cysteine). This variant aromatic placement (indicated by arrows in Figure 1) is conserved among even-numbered domains in the ZFY-related gene family (Mardon & Page, 1989; Mardon et al., 1989, 1990; Mitchell et al., 1989; Schneider-Gadicke et al., 1989; Nagamine et al., 1990; DiLella et al., 1990).

We have previously explored the structural role of Y10 in ZFY-6P by mutagenesis and characterization of relative thermodynamic stability as a function of pH or denaturants (Weiss & Keutmann, 1990). The substitutions Y10S and Y10K were observed to destabilize the finger; remarkably, stability was restored by the second-site mutation S12F. The stability of the aromatic swap revertant may now be rationalized from the structure of ZFY-6T. The packing of Y10 in the hydrophobic core is similar to that of F12 in Xfin-31 [F10 in the numbering scheme of Lee et al. (1989a,b)]. Juxtaposition of these rings and neighboring internal residues in ZFY-6T and Xfin-31 is shown in panel B of Figure 11;

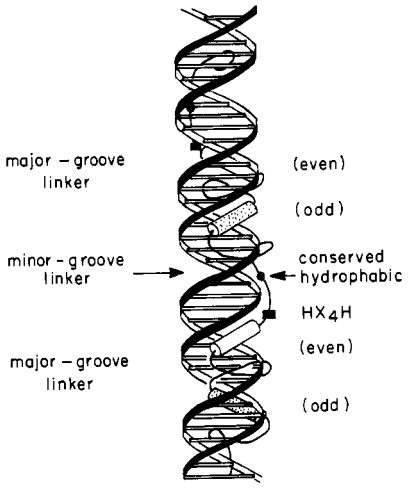

FIGURE 12: Proposed jumping-linker model of the ZFY–DNA complex exhibiting a two-finger structural repeat. This repeat is defined by alternating linker orientations: odd-numbered linkers are proposed to remain in the major groove whereas even-numbered linkers are proposed to traverse (jump) the minor groove. Such jumping is predicted to lengthen the target DNA site (>3 bp/finger) and make possible minor-groove contacts. The distinguishing features of the even linkers (i.e., histidine spacing HX_4H and the conserved hydrophobic residue) are indicated by solid rectangles and solid circles, respectively. Their positions are schematic only and not meant to convey details of the structure. Putative recognition helices of even-numbered domains are shown in the major groove as open cylinders and those of odd-numbered domains as stippled cylinders. For clarity each finger is shown to span 5–5.5 base pairs; in fact, model-building studies suggest the span to be 3–3.5 base pairs (Berg, 1990). Alternative models that exhibit a one-finger repeat [e.g., model I of Fairall et al. (1986)] are also possible. The feasibility of the "jumping-linker" model is illustrated by molecular modeling in Figure 11D.

ZFY-6T residues are shown in green and Xfin-31 residues in red. Their similar interactions suggest that alternation of aromatic sites in ZFY may not be a necessary or general feature of second-order repeats. Indeed, alternating Zn-finger sequences have been identified in *Xenopus laevis* that contain only the consensus aromatic residue (F12; Nietfeld et al., 1989).

Functional Implications of the Aromatic Position. Although structural equivalence of Y10 and F12 is apparent in the packing of the hydrophobic core, such equivalence does not account for the striking conservation of aromatic alternation among avian and mammalian ZFY-related sequences (DiLella et al., 1990). In the absence of selection pressure, genetic drift between these two packing arrangements would be expected, particularly in light of the apparently proper folding of the "aromatic-less" analogues Y10S and Y10K (Weiss & Keutmann, 1990). Accordingly, there are likely to be associated changes in the protein surface (as distinct from the hydrophobic core) that are involved in function (Kochoyan et al., 1991).

In ZFY-6T the Y10 phenolic hydroxyl group is positioned to form hydrogen bonds with either or both of the S12 and N17 side chains. The possible importance of such interactions among side chains has been demonstrated by cocrystallographic studies of prokaryotic helix–turn–helix (HTH) repressors: analogous interactions are observed to orient functional groups that are in turn involved in DNA recognition (Aggarwal et al., 1988; Jordan & Pabo, 1988). It is also possible that Y10 could itself hydrogen-bond directly with DNA, as observed for an internal tyrosine (Y22) in the λ repressor–operator cocrystal structure (Jordan & Pabo, 1988) and for a conserved tryptophan in the Engrailed homeodomain–DNA cocrystal structure (Kissinger et al., 1990). Either of these possibilities could generate functional constraints limiting variation at position 10. A more detailed understanding of the configuration of Y10 and its influence on the orientation of surface functional groups will require further refinement of the 2D NMR structure. It is intriguing that among even-numbered ZFY-related sequences residue 17 (which may interact with Y10) is conserved *within* individual fingers [e.g., asparagine in finger 6 in chicken, mouse, and man (DiLella et al., 1990)] but exhibits significant variation *between* fingers (e.g., aspartic acid in finger 8; I, V, or L in finger 10; and glycine in finger 12). This pattern of conservation *and* variation would be expected of a DNA-contact residue.

Alternation of Histidine Spacing. The even-specific HX_4H spacing between conserved histidines is associated with a distortion in the C-terminal α-helix. Such a distortion was originally predicted by Berg (1988) on the basis of molecular modeling studies and is apparently required to maintain tetrahedral coordination of Zn^{2+}. Comparison of the HX_3H and HX_4H metal-binding sites is shown in panel C of Figure 11. Departure from α-helical geometry occurs primarily near the second histidine (H26). This is reflected in the primary NMR data as an anomalous $^3J_{HN\alpha}$ coupling at residue H26 and variation in the intensity of helix-association NOEs from residues B25–B27. HX_4H spacing has previously been studied qualitatively in an ADR1-derived finger and appears to exhibit a different pattern of NOE intensities (Parraga et al., 1988). The significance and generality of these distortions are presently not clear; a possible role in DNA binding is discussed below.

(II) Jumping-Linker Model of the ZFY Protein–DNA Complex

What accounts for the striking conservation of the two-finger ZFY repeat among related genes in reptiles, birds, and mammals? DNA-binding proteins containing two-finger repeats may arise in vertebrate evolution through the duplication of an ancestral two-finger unit. In this view the fundamental repeat in the protein–DNA complex would be one finger; the apparent second-order repeat in ZFY would simply reflect a corresponding duplication in the target DNA (Page et al., 1987).

We propose as an alternative working hypothesis that the two-finger repeat reflects an essential element of the protein–DNA complex and explore in this section the implications of this hypothesis. In so doing, we note that it is likely that Zn fingers may be used in different ways in different systems; accordingly, the model proposed in this section is meant to apply to second-order (ZFY-like) repeats and not to Zn-finger proteins in general. We make three structural assumptions. Assumption 1: The structural unit of the ZFY–DNA complex is a two-finger repeat. Assumption 2: DNA–finger interactions occur in the major groove, and odd- and even-numbered fingers employ a similar mechanism of binding. Assumption 3: The two-finger repeat in ZFY reflects an alternation in linker orientation. Odd-numbered (consensus) linkers connect adjoining fingers in the major groove; even-numbered (nonconsensus) linkers traverse (jump) the minor groove.

A model incorporating these features is shown in schematic form in Figure 12. The feasibility of such a model, with linkers seven residues in length (i.e., $HX_{3,4}H–X_7–CX_{2-4}C$), is shown in panel D of Figure 11. In this model the fingertip and adjoining C-terminal α-helix are positioned in the major groove; such use of an α-helix is observed in cocrystallographic complexes of helix–turn–helix (HTH) proteins (Aggarwal et al., 1988; Jordan & Pabo, 1988; Otting et al., 1990; Kissinger et al., 1990) and proposed as a mode of Zn-finger–DNA

interaction (Fairall et al., 1986; Berg, 1988; Gibson et al., 1988). The putative recognition helices of even-numbered domains are shown as open cylinders and those of odd-numbered domains as stippled cylinders; these are oriented identically with respect to the major groove in accord with assumption 2. The odd-numbered linkers (i.e., those following odd-numbered domains) continue in the major groove, and even-numbered linkers (i.e., those following even-numbered domains) traverse the minor groove in accord with assumption 3. The even-specific HX_4H linker exit and the conserved hydrophobic residue (asterisks in Figure 1; see below) are indicated in Figure 12. On the assumption that successive fingers in the major groove contact 3 base pairs per finger (Berg, 1990), the two-finger/linker unit (assumption 1) is predicted to span 10–11 base pairs per unit, i.e., one turn of B-DNA. This estimate includes 6 base pairs per two contiguous fingers and 4–5 base pairs per even-numbered linker as it traverses the minor groove. Interestingly, Klug and colleagues also calculate 10–11 base pairs per two-finger unit, using the different assumption of 5–5.5 base pairs per finger (Fairall et al., 1986).

In the following subsections we first discuss the relationship of our model to earlier models of Zn-finger–DNA recognition and evaluate the appropriateness of the above structural assumptions. The even finger/linker submotif, HX_4HX_3–hydrophobic residue–X_3, is proposed to span the minor groove; the hydrophobic residue presumably packs against a deoxyribose moiety. Interestingly, analogous use of a hydrophobic residue in a linker that crosses the minor groove has recently been observed in a homeodomain–DNA complex (Otting et al., 1990; Kissinger et al., 1990; see below). We then extend the model—on the basis of certain sequence similarities—to a portion of the TFIIIA–DNA complex (fingers 1–4). Remarkably, our revised interpretation of the TFIIIA–DNA complex is supported by a reinterpretation of recent high-resolution hydroxyl radical footprinting (Churchill et al., 1990). The reinterpretation presented below differs in part from that given in the original paper.

Previous Models of Zn-Finger–DNA Recognition. Klug and colleagues orginally proposed two general models of the Zn-finger protein–DNA complex (Fairall et al., 1986). In model I the protein follows the helical path of the major groove; consecutive fingers are equivalently oriented with respect to the DNA, and so the structural repeat is one finger. In model II the protein lies along one face of the double helix; alternating Zn fingers make inequivalent DNA contacts in accord with a two-finger repeat (Berg, 1987). Although this model predicts the existence of distinct odd- and even-specific architectures, widely separated N- and C-termini of the Zn finger—as observed in the 2D NMR structures determined to data—make model II unlikely (Berg, 1988, 1990). Klug and colleagues have recently proposed in schematic form a revised version of model II (Churchill et al., 1990), which is meant to incorporate the 2D NMR structure of Xfin-31 (Lee et al., 1989a,b). Alternate linkers traverse the minor groove essentially *perpendicularly* to the double-helical axis. On the basis of model-building experiments, it is not clear that this schematic model can be physically realized with linkers of seven residues.

The jumping-linker model retains the assumption of Churchill et al. (1990) that every other linker traverses the minor groove (assumption 3; see above). However, the latter model differs from that presented here in the assumption by Churchill et al. (1990) of 5–5.5 base pairs per finger and in the absence of "skipped" regions of the major groove. In the ZFY model (Figure 12) the minor groove is traversed approximately *parallel* to the helical axis; this aligns successive even–odd fingers on one face of the DNA, skipping a portion of the intervening major groove. The skipped major groove is located on the "back" surface of the DNA in an arrangement reminiscent of the prokaryotic helix–turn–helix repressors [reviewed by Aggarwal and Harrison (1990)]. Fortuitously, the prediction of skipped regions of the major groove in association with HX_4HX_3–hydrophobic finger/linker units may be tested in TFIIIA since the third finger of TFIIIA exhibits such a sequence (Miller et al., 1985). The TFIIIA–DNA complex is discussed below.

Appropriateness of Assumption 2: Similar Major-Groove Interactions of Odd and Even Zn Fingers. The general similarity of ZFY-6T and previously characterized classical Zn fingers (part II under Results) suggests that odd and even ZFY fingers share a *common local mechanism* of DNA binding. Such binding is likely to occur in the DNA major groove and is proposed to be mediated by the fingertip and adjoining C-terminal α-helix (Berg, 1990). Consistent with this hypothesis, the sequence of odd- and even-numbered ZFY domains predicts a similar array of functional groups along this surface (Figure 1). If this view is correct, then although odd and even fingers exhibit similar *local* contacts with DNA, they may nonetheless play different *global* roles in the complex. In the model shown in Figures 11D and 12 the key difference lies in the orientation of their respective *linkers*.

Appropriateness of Assumption 3: Alternating Linker Motifs. Among the general class of Zn-finger proteins variability is observed in the length and sequence of the linkers. The majority of linkers contain seven residues (Gibson et al., 1988). The N-terminal one or two residues may extend the α-helix of the preceding finger, and the C-terminal three residues form part of the β-hairpin of the next finger. 2D NMR studies of a two-finger peptide from SWI5 indicate that successive fingers are flexibly linked (Neuhaus et al., 1990). Nonconsensus features of particular linkers, such as the foreshortened sixth and seventh linkers in TFIIIA (Brown et al., 1985), are proposed to relate to local features of the protein–DNA complex, such as DNA bending (Vrana et al., 1988; Berg, 1990).

ZFY-related Zn-finger proteins exhibit an alternation in the sequences of odd- and even-numbered linkers as well as fingers (Figure 1). Although both sets of linkers contain seven residues, systematic differences are observed in the placement of hydrophobic and hydrophilic residues; these suggest an altered spatial relationship to DNA. Linker sequences from avian, murine, and human genes are shown in Table II. [In this table we focus on linkers 4–13, since domain 3 is not conserved as a metal-binding site and may function as a spacing element in the protein (Weiss et al., 1990); the first two fingers and linkers are also less well conserved (DiLella et al., 1990)]. Remarkably, a conserved hydrophobic residue is observed in the center of each even-numbered linker; such a hydrophobic residue is not observed among the odd-numbered linkers (which contain lysine at this position; arrow in Table II) nor among Zn-finger proteins generally (Gibson et al., 1988). Because hydrophobic residues are ordinarily conserved as part of a hydrophobic core, its presence in a presumably flexible region (Neuhaus et al., 1990) seems paradoxical and suggests the existence of functional constraints in the protein–DNA complex. A similar "paradox" occurs in the N-terminal linker of homeodomains (i.e., conservation of a hydrophobic residue in a flexible region). The role of this hydrophobic residue has recently been defined by 3D NMR and crystallographic analyses of the specific homeodomain–DNA complex. We will briefly review the homeodomain paradox and suggest an analogous role for the conserved hy-

Table II: Linker Sequences among ZFY-Related Genes[a]

	Linker 1-2 ⇓	Linker 2-3
Chicken	PEHLLTKKKYR	KLTNKTEKLIE
human ZFY	PEH-LAKKKYH	KLTSKAEKAIE
human ZFX	PEH-LAKKKYR	KLTSKAEKAIE
m. mus Zfy-1	PEY-LANKKYH	KLTIKTEKTTE
m. mus Zfy-2	PEY-LANKKYH	KLTIKTEKTTE
	Linker 3-4 ⇓	**Linker 4-5**
Chicken	REKGVNKMHK	SKNEPHI
human ZFY	KEKGANKMHK	SKNEPHI
human ZFX	KEKGANKMHK	SKNEPHI
m. mus Zfy-1	TEKGVNKTCK	RKKEPHI
m. mus Zfy-2	TEKGVNKTCK	RKKEPHI
	Linker 5-6 ⇓	**Linker 6-7**
Chicken	TGEKPYQ	SKETSSK
human ZFY	TGEKPYQ	SKEMPFK
human ZFX	TGEKPYQ	SKEMPFK
m. mus Zfy-1	TGEKPYE	SKEIPLK
m. mus Zfy-2	TGEKPYE	SKEIPLK
	Linker 7-8 ⇓	**Linker 8-9**
Chicken	QESKTHQ	TKDYPHK
human ZFY	QESKTHQ	TKDYPHK
human ZFX	QESKTHQ	TKDYPHK
m. mus Zfy-1	QESRTHQ	TKAYPHK
m. mus Zfy-2	QESRTHQ	TKAYPHK
	Linker 9-10 ⇓	**Linker 10-11**
Chicken	KGKKLHQ	TKDLPFR
human ZFY	KGKKMHQ	TKDLPFR
human ZFX	KGKKMHQ	TKDLPFR
m. mus Zfy-1	KSKKMHQ	TKNVPFK
m. mus Zfy-2	KSKKMHQ	TKNVPFK
	Linker 11-12 ⇓	**Linker 12-13**
Chicken	SGRKVYQ	TKDYPHR
human ZFY	SGRKVYQ	TKDYPHR
human ZFX	SGRKVYQ	TKDYPHR
m. mus Zfy-1	SSRKVYQ	TKDYPHS
m. mus Zfy-2	SSRKVYQ	TKDYPHR

[a] Sequences shown were obtained from DiLella et al. (1990). The arrow indicates a conserved lysine in position 4 of the odd-numbered linkers; the corresponding position in the even-numbered linkers is conserved as a hydrophobic residue (boldface and underlined).

drophobic residue in even-numbered linkers in ZFY.

Analogy to Homeodomains. The homeodomain consists of three α-helixes, as shown in panel A of Figure 13; helices 2 and 3 form a helix–turn–helix unit. [In the original 2D NMR structure of the Antennapedia domain the third helix has a kink; the C-terminal section was accordingly designated helix 4 (Qian et al., 1989). This kink is not observed in the Engrailed protein–DNA complex (Kissinger et al., 1990) and so is not shown in Figure 13.] Residues 0–6 [using the numbering

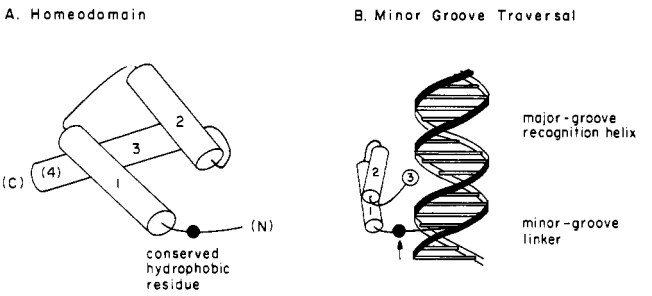

FIGURE 13: Analogy between homeodomain- and ZFY-related jumping linkers. (A) Cylinder representation of the three α-helices in the homeodomain as determined by 2D NMR (Antennapedia; Qian et al., 1989) and X-ray crystallography (Engrailed; Kissinger et al., 1990). The C-terminal extension of helix 3 is designated by Qian et al. (1989) as helix 4. The N-terminal arms of these domains (which presumably are linkers in the intact proteins) contain a hydrophobic residue as indicated; the hydrophobic character of this residue is conserved among homeodomains (see Table III). Helices 2 and 3 form a helix–turn–helix unit. (B) Structure of the specific homeodomain–DNA complex determined by X-ray crystallography (Kissinger et al., 1990) and 3D NMR (Otting et al., 1990). The recognition helix (helix 3, shown end-on as a circle rather than a cylinder in this orientation) contacts DNA in the major groove. A novel feature of this complex is the traversal of the minor groove by the N-terminal linker. The conserved hydrophobic residue (arrow) contacts a deoxyribose moiety as it exists from the major groove to the minor groove. Analogous use of a hydrophobic residue is proposed in the ZFY–DNA complex (Figure 12).

scheme of Qian et al. (1989)] were observed to be disordered in solution, and a conserved hydrophobic residue at position 8 (tyrosine in the Antennapedia domain; and phenylalanine in Engrailed; see Table III) made no significant nonlocal interactions. Since homeodomains are typically imbedded in larger proteins, this region is more appropriately considered a flexible *linker* between discrete domains (rather than an N-terminal *arm*, as in the protein fragments used for structural analysis). The homeodomain contains a well-ordered hydrophobic core (Qian et al., 1989; Kissinger et al., 1990), which involves a set of residues that are strictly conserved among eukaryotic sequences (Scott et al., 1989). Tyr8 is the only hydrophobic residue whose conservation is not easily rationalized from the 2D NMR structure of the protein in solution.

The structure of a complex between a homeodomain and a specific DNA site has recently been described in solution (Otting et al., 1990) and in the crystal state (Kissinger et al., 1990). In these structures the N-terminal linker (KRPRTAFS in Engrailed; the conserved hydrophobic residue is indicated in boldface) traverses from the major groove to the minor groove; novel arm-related base and backbone contacts are observed in the minor groove and involved residues conserved among homeodomains (Scott et al., 1989). Surprisingly, groove crossing also involves the highly conserved hydrophobic residue in the linker, which is observed to pack agaisnt a deoxyribose moiety. This interaction is shown in schematic form in panel B of Figure 13. Putative groove-crossing sequences in representative homeodomains are given in Table III. Pabo and colleagues speculate that such groove-crossing sequences may play an important role in enabling contiguous protein domains to bind to adjacent DNA sites (as in the bipartite POU motif) or in establishing protein–protein contacts with neighboring DNA-binding factors (Kissinger et al., 1990).

We propose that the conserved hydrophobic residue in even-numbered ZFY linkers plays an analogous role in the protein–DNA complex; i.e., it participates in minor-groove traversal (jumping) by packing against a deoxyribose moiety (Figures 11D and 12). We further propose that the even-

Table III: Groove-Traversal Sequences in Homeodomains[a]

Engr[b]	E K R P R T A **E** S [helix 1]	hox-1.5	S K R G R T A **Y** T
Antp[c]	R K R G R Q T **Y** T [helix 1]	hox-2.4	R R R G R Q T **Y** S
dfd	P K R Q R T A **Y** T	unc-86	K K R K R T S **I** A
lab	N N S G R T N **E** T	Oct-1	R R K K R T S **I** E
AbdB	V R K K R K P **Y** S	Oct-2	R R K K R T S **I** E
ftz	S K R G R Q T **Y** T	Oct-3	R K R K R T S **I** E
Ubx	R R R G R Q T **Y** T	Ch. Oct	R R K K R T S **I** E
en	E K R P R T A **E** S	Pit-1	K R K R R T T **I** S
eve	V R R Y R T A **E** T	PHO2	Q R P K R T R **A** K
prd	Q R R C R T T **E** S	mec-3	R R G P R T T **I** K
cut	S K K Q R V L **E** S	lin-11	R R G P R T T **I** K
al	S P K G K S S **I** S	Isl-1	T T R V R T V **L** N
α2	K P Y R G H R **E** T		

[a] Sequences shown were obtained from Scott et al. (1989). [b] The structure of the Engrailed protein–DNA complex has been described by Pabo and co-workers (Kissinger et al., 1990). [c] The NMR structure of the Antp domain and its DNA complex has been described by Wuthrich and co-workers (Qian et al., 1989; Otting et al., 1990).

specific HX_4H histidine spacing is designed to permit exit of the linker from the DNA major groove. The "jumping linker" in turn provides a structural mechanism (i) to lengthen the target DNA site to be recognized and (ii) to make possible protein–DNA interactions in the minor groove. Testing these predictions will require a combination of biochemical characterization of mutant Zn-finger proteins and structural analysis of representive Zn-finger–DNA complexes (by X-ray crystallography or 3D NMR).

Analogy to TFIIIA. The model of the ZFY–DNA complex proposed above is supported in part by a reinterpretation of the quantitative analysis of hydroxyl radical footprinting of the TFIIIA–DNA complex (Churchill et al., 1990). Unlike ZFY, there is no simple predominant second-order or higher order repeat among the nine TFIIIA fingers or linkers (Miller et al., 1985). In fact, detailed inspection of the TFIIIA hydroxyl radical footprint does not provide compelling evidence for any one single or simple mode of binding. The absence of a clear pattern may reflect local variations due to the differences in finger and linker sequences, as discussed by Churchill et al. (1990). Berg (1990) has proposed, for example, that foreshortening of the linkers preceding and following finger 6 is associated with a kink in the target DNA site; such variation in the three-dimensional structure of the complex would necessarily complicate straightforward interpretation of a "two-dimensional" footprint. TFIIIA may also bind DNA with more than one mechanism, in which case the observed footprint would reflect more than one underlying pattern.

Our interest in TFIIIA fingers is motivated by a striking similarity between TFIIIA finger 3 and its adjoining C-terminal linker to the even ZFY motif: each shares the HX_4HX_3–hydrophobic residue pattern discussed above. The amino acid sequences of fingers 1–4 of TFIIIA are shown in panel A of Figure 14; the first four domains may be classified (in ZFY terms) as odd–odd–even–odd. Do footprinting studies of the TFIIIA–DNA complex support a ZFY-type model in this region of the protein? We predict that fingers 1–3 are positioned contiguously in the major groove and span one turn of the double helix (i.e., three fingers per DNA turn and 3–3.5 base pairs per finger). We further predict that the linker between fingers 3 and 4 traverses the minor groove approximately parallel to the helical axis, positioning finger 4 in the major groove on the same face as finger 3. These predictions are illustrated in schematic form in panel D of Figure 14. Remarkably, they are supported by a reinterpretation of data previously in the literature.

The overall arrangement of TFIIIA Zn fingers with respect to its DNA target site has been investigated by comparative footprint analysis of deletion mutants (Vrana et al., 1986). Such studies indicate that fingers 1–4 bind approximately between base pairs 98 (finger 1) and 70 (finger 4). Accordingly, we focus on this region of the TFIIIA high-resolution hydroxyl radical footprint (Churchill et al., 1990), which is replotted in modified form in panel B of Figure 14. The hydroxyl radical technique evaluates the extent to which protein binding reduces the rate of DNA backbone cleavage by hydroxyl radical ($OH^{\bullet}$) (Tullius, 1989); results may be illustrated as a contour plot representing levels of protection in a cut-cylinder model of DNA (Churchill et al., 1990). In accord with a ZFY-like model, a continuous region of protection is observed for one turn of helix (between base pairs 90 and 80 of the coding strand), which then traverses the minor groove. The upstream protected region, which we ascribe to finger 4, then continues along the same face of the double helix. Additional evidence for an interruption in major-groove binding is provided by the pattern of protection (and accessibility) of guanine N7 sites in the major groove from methylation by dimethyl sulfate (DMS) (Fairall et al., 1986); protected sites are indicated in panel C of Figure 14 by closed circles and accessible sites by open circles. Interestingly, guanines at positions 75 (noncoding strand) and 78 (coding strand) are unprotected (indicated by arrows in panel C of Figure 14), further suggesting that TFIIIA skips this portion of the major groove. The protection pattern upstream of this region (fingers 5–9; 5S ICS DNA base pairs 70–40) is not readily interpreted and is likely to reflect distortions in linker length or DNA geometry, as emphasized by Berg (1990). Since the DNA target site for ZFY (or related proteins) has not been identified, corresponding footprinting studies of the specific ZFY–DNA complex cannot presently be performed.

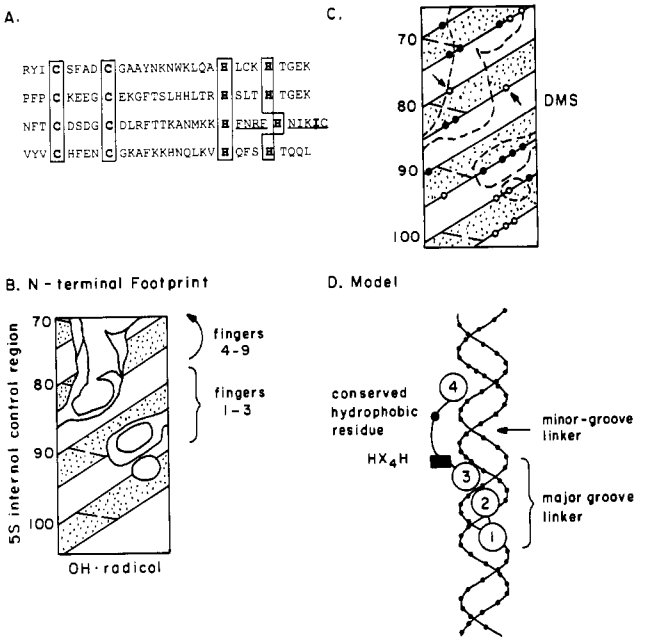

FIGURE 14: Analogy between TFIIIA and ZFY. (A) Sequence of fingers 1–4 of TFIIIA (Miller et al., 1985). The third finger is notable for a HX_4H inter-histidine spacing (underlined) followed by a variant linker containing a hydrophobic residue (shown in boldface) at the same position as the even-numbered linkers of ZFY. (B) Portion of the high-resolution hydroxyl radical ($OH^{\bullet}$) footprint (bases 70–100) of the TFIIIA–ICS DNA (5S RNA gene) recently reported by Churchill et al. (1990). This region is thought to be contacted by the N-terminal five fingers of TFIIIA (Vrana et al., 1988) and is remarkable for a minor-groove traversal pattern following one turn of the DNA double helix. The footprint (replotted in modified form from the original paper) is shown as a contour plot in a cut-cylinder representation of DNA; levels 4, 5, and 6–8 are shown. The minor-groove surface is stippled. (C) Corresponding portion of the TFIIIA–DNA footprint showing the pattern of protection (and accessibility) of guanine N7 sites in the major groove from methylation by dimethyl sulfate (DMS). These data are replotted in modified form from Fairall et al. (1986) and correspond to panel B; for clarity, one contour level of the hydroxyl radical footprint is shown as a dotted line. Protected guanine N7 sites are represented by close circles and accessible ones by open circles. Arrows indicate two unprotected sites (positions 75 and 78) whose accessibility is proposed to reflect minor-groove crossing by the linker between fingers 3 and 4. (D) Proposed model of the N-terminal portion of the TFIIIA–DNA complex. In this model fingers 1–3 bind successively in the major groove (spanning one turn) and are connected by major-groove linkers. Finger 4 is connected across the minor groove by a jumping linker and binds in the major groove on the same face of the DNA as finger 3. The (finger 2–consensus linker–finger 3–variant linker–finger 4) domains of TFIIIA are thus proposed to resemble an odd–even–odd unit of ZFY (see Figures 11D and 12). As in Figure 12, the distinguishing features of the variant linker (i.e., histidine spacing HX_4H and the conserved hydrophobic residue) are schematically indicated by a solid rectangle and a solid circle, respectively.

Conclusions

The two-finger repeat observed in the ZFY-related gene family exhibits systematic alternation between odd (consensus) and even (nonconsensus) Zn fingers and linkers. As a first step in the structural characterization of the ZFY two-finger repeat, we have determined the solution structure of a representative even domain. The similarity of this structure to consensus-type Zn fingers suggests that odd- and even-numbered ZFY domains bind DNA by a common local mechanism. This similarity includes (i) an N-terminal β-hairpin/C-terminal α-helix folding unit, (ii) a hydrophobic core defined by a conserved framework of hydrophobic and aromatic residues, and (iii) an apparent equivalence between internal packing interactions at alternative aromatic sites Y10 and F12. The latter result rationalizes the successful design of an aromatic swap revertant as previously described (Weiss & Keutmann, 1990). Structural differences between ZFY-6T and consensus-type Zn fingers are observed in the α-helix/linker "exit" related to the HX_4H metal-binding site. The variant HX_4HX_3–hydrophobic finger/linker submotif is predicted to facilitate traversal of the minor groove by adjoining nonconsensus (even-numbered) linkers in the second-order ZFY repeat. This model (designated the jumping-linker model) is supported in part by a reinterpretation of the high-resolution hydroxyl radical footprinting of the TFIIIA–DNA complex (fingers 1–4). The proposed use of a hydrophobic residue in a linker that crosses the minor groove is analogous to a similar interaction in the recently described cocrystal structure of a homeodomain–DNA complex (Kissinger et al., 1990).

Added in Proof

A high-resolution NMR structure of an isolated finger from a human enhancer-binding protein has recently been described by Gronenborn and colleagues (Omichinski et al., 1990). This structure also exhibits the $\beta\beta\alpha$ motif and is notable for a distortion in α-helical geometry associated with variant ligand spacing HX_5H. This distortion is distinct from that described here in association with the HX_4H sequence pattern. Alternative ligand spacings $HX_{3-5}H$ thus appear to encode an independent structural feature of the classical Zn finger motif.

Acknowledgments

We thank K. A. Mason for technical assistance; A. Bax for generously providing advice and NMR time in the early stages of the project; M. Karplus for computer time; C. Kissinger, C. Pabo, D. C. Page, P. A. Sharp, and A. G. Redfield for helpful discussion and communication of results prior to publication; J. Lee, B. Stockman, and M. Westler for advice regarding NMR methods; D. Case and P. E. Wright for the coordinates of Xfin-31 and helpful discussion; G. M. Clore and A. Gronenborn for the coordinates of an enhancer-binding finger; and C. T. Walsh for a critical reading of the manuscript. The NMR Facility at Harvard Medical School is supported by an NIH Shared Instrumentation Grant (1 S10 RR04862-01) and an award from the National Health Resources Foundation. The NMR Resource at the University of Wisconsin at Madison is supported by NIH Grant RR-02301 and NSF Grant PCM-8415048. The NMR Resource at the MIT Francis Bitter National Magnetic Laboratory is supported by NIH Grant RR-00995.

Supplementary Material Available

Two tables listing NOE restraints used in distance-geometry calculations and stereospecific assignments and two figures showing a gel filtration chromatogram of the ZFY-6T/Zn^{2+} complex and possible coordination schemes of Zn^{2+} with two Cys and two His ligands (11 pages). Ordering information is given on any current masthead page.

Registry No. L-Cys, 52-90-4; L-His, 71-00-1; Zn, 7440-66-6.

References

Aggarwal, A. K., & Harrison, S. C. (1990) *Annu. Rev. Biochem. 59*, 933–969.

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., & Harrison, S. C. (1988) *Science 242*, 899–907.

Barany, G., & Merrifield, R. B. (1979) in *The Peptides* (Gross, E., & Meienhofer, J., Eds.) Vol. 2, pp 1–284, Ac-

ademic Press, New York, NY.
Berg, J. M. (1987) *Science 232*, 485–488.
Berg, J. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 99–487.
Berg, J. M. (1990) *Annu. Rev. Biophys. Biophys. Chem. 19*, 405–421.
Brown, R. S., Sanders, C., & Argos, S. (1985) *FEBS Lett. 186*, 271–274.
Churchill, M. E. A., Tullius, T. D., & Klug, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5528–5532.
Crippen, G. M. (1981) *Distance Geometry and Conformational Calculations*, Research Studies Press, Taunton, U.K.
Crippen, G. M., & Havel, T. F. (1988) *Distance Geometry and Conformational Calculations*, Research Studies Press, Taunton, U.K.
de la Chapelle, A. (1972) *Am. J. Hum. Genet. 24*, 71–105.
DiLella, A. G., Page, D. C., & Smith, R. G. (1990) *New Biol. 2*, 49–55.
Evans, R. M., & Hollenberg, S. M. (1988) *Cell 52*, 1–3.
Fairall, L., Rhodes, D., & Klug, A. (1986) *J. Mol. Biol. 192*, 577–591.
Frankel, A. D., & Pabo, C. O. (1988) *Cell 55*, 675–676.
Frankel, A. D., Berg, J. M., & Pabo, C. O. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4841–4845.
Gibson, T. J., Postma, J. P. M., Brown, R. S., & Argos, P. (1988) *Protein Eng. 2*, 209–218.
Havel, T. F. (1990) *Biopolymers 29*, 1565–1585.
Havel, T. F. (1991) in *Progress in Biophysics and Molecular Biology* (Noble, D., & Blundell, T. L., Eds.) Permagon Press, Oxford, U.K. (in press).
Havel, T. F., & Wuthrich, K. (1984) *Bull. Math. Biol. 46*, 673–698.
Havel, T. F., Kuntz, I. D., & Crippen, G. M. (1983) *Bull. Math. Biol. 45*, 665–720.
Jandu, S. K., Ray, S., Brooks, L., & Leatherbarrow, R. J. (1990) *Biochemistry 29*, 6264–6269.
Johnson, C. E., & Bovey, F. A. (1958) *J. Chem. Phys. 29*, 1012–1018.
Jordan, A., & Pabo, C. O. (1988) *Science 242*, 895–899.
Karplus, M. (1959) *J. Chem. Phys. 30*, 11–15.
Kissinger, C. R., Liu, B., Martin-Blanc, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579–590.
Klevit, R. E., Herriol, J. R., & Horvath, S. (1990) *Proteins 7*, 214–226.
Klug, A., & Rhodes, D. (1987) *Trends Biochem. Sci. 12*, 464–468.
Kochoyan, M., Keutmann, H. T., & Weiss, M. A. (1991) *Biochemistry* (submitted for publication).
Koopman, P., Gubbay, J., Colignon, J., & Lovell-Badge, R. (1989) *Nature 342*, 940–942.
Lee, M. S., Cavanagh, J., & Wright, P. E. (1989a) *FEBS Lett. 254*, 159–164.
Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., & Wright, P. E. (1989b) *Science 245*, 635–637.
Mardon, G., & Page, D. C. (1989) *Cell 56*, 765–770.
Mardon, G., Mosher, R., Disteche, C. M., Nishioka, Y., McLaren, A., & Page, D. C. (1989) *Science 243*, 78–243.
Mardon, G., Luoh, S.-W., Simpson, E. M., Gill, G., Brown, L. G., & Page, D. C. (1990) *Mol. Cell. Biol. 10*, 681–688.
Miller, J., MacLachlan, A. D., & Klug, A. (1985) *EMBO J. 4*, 1609–1614.
Mitchell, M., Simon, D., Affara, N., Fergusson-Smith, M., Avner, P., & Bishop, C. (1989) *Genetics 121*, 803–809.
Nagamine, C. M., Chan, K., Hake, L. E., & Lau, Y.-F. C. (1990) *Genes Dev. 4*, 63–74.
Neuhaus, D., Nakaseko, Y., Nagai, K., & Klug, A. (1990) *FEBS Lett. 262*, 179–184.
Nietfeld, W., El-Baradi, T., Mentzel, H., Pieler, T., Koster, M., Poting, A., & Knochel, W. (1989) *J. Mol. Biol. 208*, 639–659.
Ominchinski, J. G., Clare, G. M., Apella, E., Sakaguchi, K., & Gronenborn, A. M. (1990) *Biochemistry 29*, 9324–9334.
Otting, G., Qian, Y., Muller, M., Affolter, M., Gehring, W., & Wuthrich, K. (1988) *EMBO J. 7*, 4305–4309.
Otting, G., Qian, Y., Billeter, M., Muller, M., Affolter, M., Gehring, W., & Wuthrich, K. (1990) *EMBO J. 9*, 3085–3092.
Page, D. C., Mosher, R., Simpson, E., Fisher, E. M. C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A., & Brown, L. G. (1987) *Cell 51*, 1091–1104.
Page, D. C., Fisher, E. M. C., McGillivrany, B., & Brown, L. G. (1990) *Nature 346*, 279–281.
Palmer, M. S., Sinclair, A. H., Berta, P., Ellis, N. A., Goodfellow, P. N., Abbas, N. E., & Fellous, M. (1989) *Nature 342*, 937–939.
Pardi, A., Billeter, M., & Wuthrich, K. (1984) *J. Mol. Biol. 180*, 741.
Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., & Klevit, R. E. (1988) *Science 241*, 1489–1492.
Parraga, G., Horvath, S., Hood, L., Young, E. T., & Klevit, R. E. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 137–141.
Qian, Y. Q., Billeter, M., Otting, G., Gehring, W. J., & Wuthrich, K. (1989) *Cell 59*, 573–580.
Schneider-Gadicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R., & Page, D. C. (1989) *Cell 57*, 1247–1258.
Scott, M. P., Tamkun, J., & Hartzel, G. (1989) *Biochim. Biophys. Acta 989*, 25–48.
Stewart, J. M., & Young, J. D. (1984) *Solid Phase Peptide Synthesis*, Raven Press, New York, NY.
Tregear, G. W., van Reitschoten, J., Sauer, R. T., Niall, H. D., Keutmann, H. T., & Potts, J. T. (1977) *Biochemistry 16*, 2817–2823.
Tullius, T. D. (1989) *Annu. Rev. Biophys. Biophys. Chem. 18*, 213–237.
Vrana, K. E., Churchill, M. E. A., Tullius, D. T., & Brown, D. D. (1988) *Mol. Cell. Biol. 8*, 1684–1696.
Wagner, G., Braun, W., Havel, F. T., Shaumann, T., Go, N., & Wuthrich, K. (1987) *J. Mol. Biol. 196*, 611–633.
Weiss, M. A., & Keutmann, H. T. (1990) *Biochemistry 29*, 9808–9813.
Weiss, M. A., Mason, K. A., Dahl, C. E., & Keutmann, H. T. (1990) *Biochemistry 29*, 5660–5664.
Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.