Alternating Zinc Fingers in the Human Male Associated Protein ZFY: Refinement of the NMR Structure of an Even Finger by Selective Deuterium Labeling and Implications for DNA Recognition†,‡

Michel Kochoyan,§,‖ 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 February 20, 1991; Revised Manuscript Received May 6, 1991

ABSTRACT: ZFY, a male-associated Zn-finger protein encoded by the human Y chromosome, exhibits a distinctive two-finger repeat: whereas odd-numbered domains fit a general consensus, even-numbered domains exhibit systematic differences. Do these odd and even sequences encode structurally distinct surfaces for DNA recognition? As a first step toward answering this question, we have recently described the sequential ^{1}H NMR assignment of a representative nonconsensus Zn finger (designated ZFY-6T) based on 2D NMR studies of a 30-residue peptide [Kochoyan, M., Havel, T. F., Nguyen, D. T., Dahl, C. E., Keutmann, H. T., & Weiss, M. A. (1991) *Biochemistry 30*, 3371–3386]. Initial structural modeling by distance geometry/simulated annealing (DG/SA) demonstrated that this peptide retained the N-terminal β-hairpin and C-terminal α-helix ($\beta\beta\alpha$ motif) observed in consensus Zn fingers. However, the precision of this initial structure was limited by resonance overlap, which led to ambiguities in the assignment of key NOEs in the hydrophobic core. In this paper these ambiguities are resolved by selective deuterium labeling, enabling a refined structure to be calculated by DG/SA and restrained molecular dynamics. These calculations provide a detailed view of the hydrophobic core and protein surface, which are analyzed in reference to previously characterized Zn fingers. Variant (even) and consensus (odd) aromatic residues Y10 and F12, shown in an "aromatic swap" analogue to provide equivalent contributions to the hydrophobic core [Weiss, M. A., & Keutmann, H. T. (1990) *Biochemistry 29*, 9808–9813], nevertheless exhibit striking differences in packing interactions: Y10—but not F12—contributes to a contiguous region of the protein surface defined by putative specificity-determining residues. Alternating surface architectures may have implications for the mechanism of DNA recognition by the ZFY two-finger repeat.

The classical CC/HH Zn-finger motif defines a highly conserved class of eukaryotic nucleic acid binding proteins involved in the regulation of gene expression (Klug & Rhodes, 1987; Evans & Hollenberg, 1988). As a first step toward defining architectural rules for sequence-dependent variations in finger structure, we have recently characterized representative Zn-finger domains from the human male associated protein ZFY (Weiss et al., 1990; Weiss & Keutmann, 1990; Kochoyan et al., 1991). This putative transcription factor, originally identified from studies of apparent sex reversal in man (de la Chapelle, 1972), is encoded by the sex-determining region of the Y chromosome (Page et al., 1987). Although its precise function is unclear, ZFY is proposed to participate in spermatogenesis (Palmer et al., 1989; Koopman et al., 1989) and may also play an accessory role in the development of the male embryo (Page et al., 1990). The ZFY-related gene family is remarkable for a conserved two-finger repeat, defined by an alternating pattern of even and odd finger/linker sequences (Page et al., 1987). Whereas odd-numbered fingers and linkers resemble a Zn-finger consensus pattern (Gibson et al., 1988), even-numbered domains exhibit systematic differences. This sequence alternation is conserved among vertebrates (DiLella et al., 1990) and is proposed in this and other systems to constitute a structural repeat in the protein–DNA complex (Nietfeld et al., 1989; Churchill et al., 1990; Kochoyan et al., 1991). The present study addresses the question: do the odd- and even-specific sequence patterns encode structurally distinct surfaces for DNA recognition?

We have recently described the 2D NMR[1] resonance assignment and preliminary structural modeling of a representative even-numbered domain (Kochoyan et al., 1991). This finger (designated ZTY-6T; Figure 1) contains an N-terminal β-hairpin and a C-terminal α-helix ($\beta\beta\alpha$ motif) similar to those of consensus-type Zn fingers (Parraga et al., 1988; Lee et al., 1989; Klevit et al., 1990; Omichinski et al., 1990). Long-range NOEs provide evidence of a hydrophobic core, whose stability is indicated by the presence of slowly exchanging ^{1}H NMR amide resonances in D_2O (Kochoyan et al., 1991) and by

† Supported by the Lucille Markey Charitable Trust and by grants from the National Institutes of Health, American Cancer Society, and 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. M.K. is supported in part by a Fondation pour la Recherche Medicale postdoctoral fellowship.

‡ The coordinates of the 15 DG structures and a table of NMR-based restraints will be deposited in the Brookhaven Data Bank upon publication.

* Address correspondence to this author at the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School.

§ Harvard Medical School.

‖ Permanent address: Biophysique Polytechnique, CNRS, URA 1254, 91128 Palaiseau, France.

⊥ Massachusetts General Hospital.

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

0006-2960/91/0430-7063$02.50/0

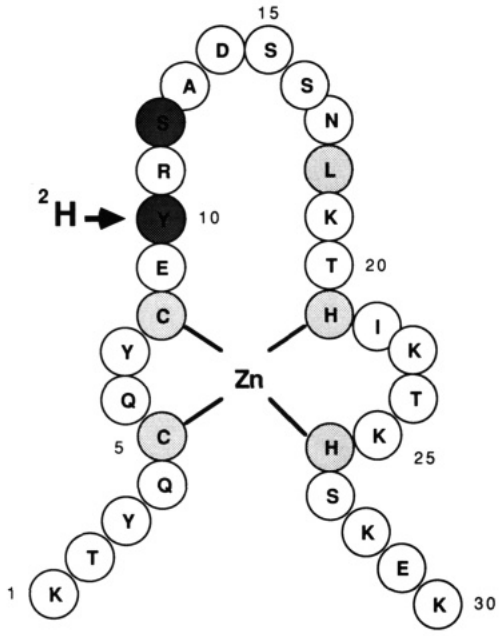

FIGURE 1: Schematic representation of the primary structure of ZFY-6T (derived from exon residues 162–191 in the *ZFY* gene; Page et al., 1987). The arrow indicates Y10, the site of selective ^{2}H labeling. Conserved residues involved in metal coordination (C5, C8, H21, and H26) are lightly shaded. Residues 10 and 12, which exhibit an aromatic swap between even and odd domains (Weiss & Keutmann, 1990), are heavily shaded.

complementary optical studies of guanidine- and pH-induced denaturation (Weiss & Keutmann, 1990). Nevertheless, the precision of the initial distance geometry/simulated annealing (DG/SA) structure was limited by degeneracy of the ring resonances of the central aromatic residue (Y10; Figure 1). Key long-range NOEs in the hydrophobic core, ordinarily expected to provide important restraints in such calculations, were of limited usefulness in the DG/SA reconstruction due to the large distance-bound correction required by the experimental ambiguity. To circumvent this limitation, we describe in this paper the use of selective ^{2}H labeling to distinguish ortho- and meta-specific NOEs involving the central aromatic ring. A more precise DG/SA model is calculated and refined by restrained molecular dynamics (Clore et al., 1985). The resulting ensemble of structures provides a more detailed model of the hydrophobic core and protein surface. This model is analyzed in reference to even-specific sequence features and compared to previously characterized CC/HH Zn-finger structures (Lee et al., 1989; Klevit et al., 1990; Omichinski et al., 1990). The variant aromatic group Y10 is shown to contribute to the protein surface differently than either residue 10 or residue 12 in consensus Zn fingers. Remarkably, this change in surface architecture adjoins putative specificity-determining residues in or near the DNA recognition helix, as determined by genetic (Nardelli et al., 1991) and X-ray crystallographic studies (Pavletich & Pabo, 1991) of other Zn-finger systems.

Materials and Methods

Peptide Synthesis and Characterization. ZFY-6T (Figure 1) was synthesized by the solid-phase procedure (Barany & Merrifield, 1979; Stewart & Young, 1984) and purified following reduction by reverse-phase HPLC as previously described (Weiss et al., 1990; Kochoyan et al., 1991). Purity was evaluated by analytical HPLC, composition, and sequence analysis of preview (Tregear et al., 1977) as previously described (Weiss et al., 1990).

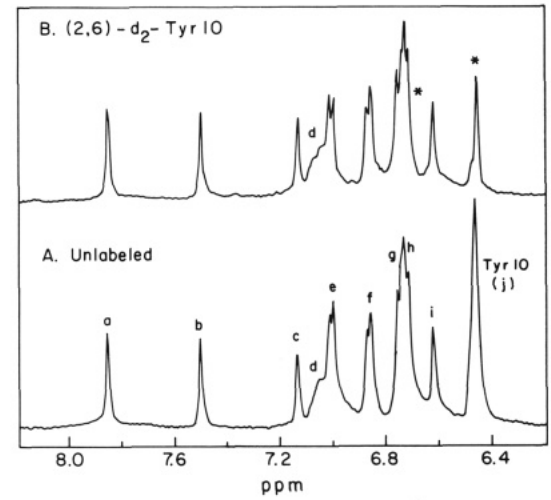

FIGURE 2: (A) One-dimensional ^{1}H NMR spectrum of the ZFY-6T/Zn^{2+} complex at 25 °C and pD 6.0 (direct meter reading). Slowly exchanging amide resonances have previously been exchanged to D_2O. (B) Corresponding spectrum of the labeled finger containing *o*-tyrosine-d_2 at position 10 (arrow in Figure 1). Following selective labeling, the compositive Y10 (ortho + meta) aromatic resonance at 6.4 ppm (resonance j in panel A) and the minor broad component at 6.7 ppm are reduced to approximately half of their former intensity (indicated by two asterisks in panel B). The following resonances are labeled in panel A: (a) H26-Hϵ, (b) H21-Hϵ, (c) H21-Hδ, (d) Y10 meta minor state, (e) Y3 ortho, (f) Y7 ortho, (g) Y3 meta, (h) Y7 meta, (i) H26-Hδ, and (j) Y10 major state.

Isotopic Labeling. Tyrosine-*2′,3′,5′,6′-d*$_4$ was prepared as the *t*-BOC *O*-benzyl ester by Cambridge Isotope Laboratories, Inc. (Woburn, MA), for use in automated peptide synthesis. The 3′,5′ deuterons were back-exchanged by protons (efficiency 90%) in anhydrous ^{1}HF in the course of cleavage from the resin and deprotection (reaction conditions: 0 °C and 50 min). Incorporation of the ^{2}H label was verified by ^{1}H NMR (Figures 2 and 3).

NMR Sample Preparation. The reduced peptides were dissolved in 0.7 mL of degassed NMR buffer (see below). The concentration of unlabeled peptide was ca. 5 mM and labeled peptide ca. 1 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. For NMR measurements deuterated 50 mM Tris-HCl (Merck Isotopes, Inc.) was used. To delay oxidation of the peptide, buffers and solvents were flushed with Ar or N_2 immediately prior to use.

NMR Methods. Spectra were recorded at 500 MHz at the Harvard Medical School NMR center. 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 as previously described (Kochoyan et al., 1991).

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 (<3.3 Å), or weak (<4.3 Å) on the basis of the 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

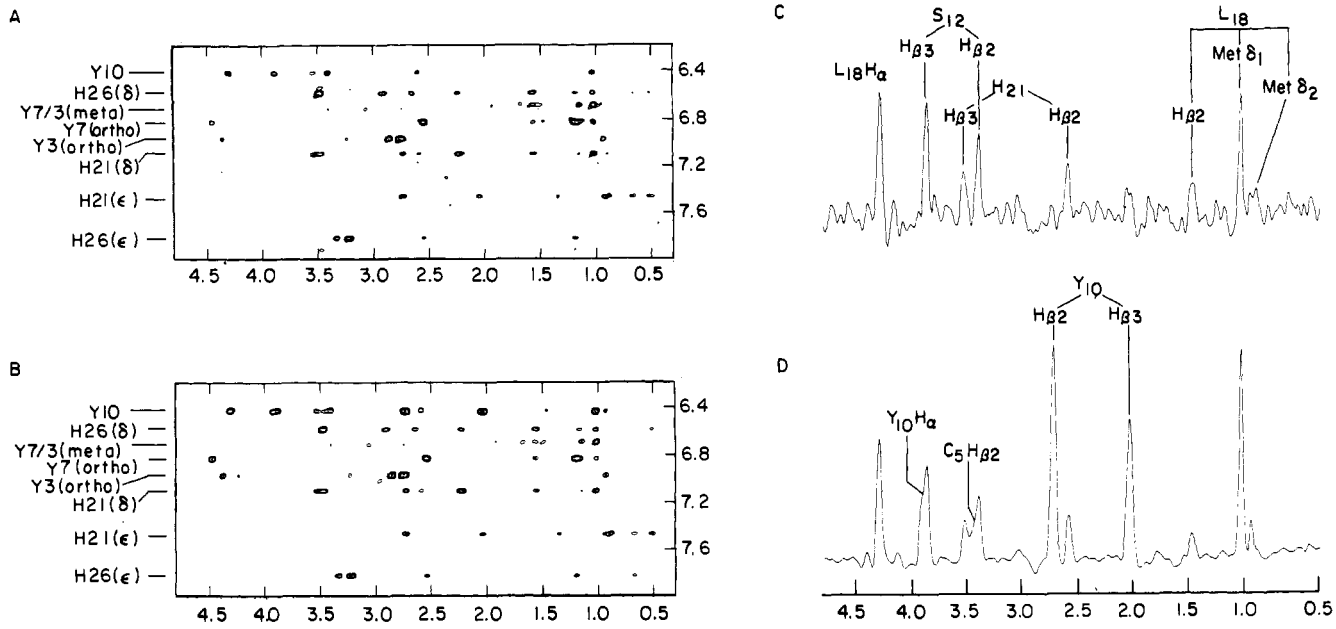

FIGURE 3: NOESY spectra of (A) labeled and (B) unlabeled Zn fingers. Contour levels are normalized relative to unlabeled sites. Slices corresponding to the Y10 resonance (6.4 ppm) are shown in panels C (labeled) and D (unlabeled). The mixing time was 200 ms; conditions were otherwise as described in the caption to Figure 2. The concentration of unlabeled protein was 5 mM; the concentration of labeled protein was 1 mM, thus accounting for the lower signal to noise in the latter spectrum. Assignments shown are as described (Kochoyan et al., 1991).

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 pathway was apparent in the initial DG model; Kochoyan et al., 1991). Distance-bound corrections were made for those methyl groups and methylene protons for which stereospecific assignments could not be obtained (Wuthrich, 1990). (ii) Dihedral angle restraints were introduced on the basis of J coupling constants in accord with the Karplus equation (Karplus, 1959) as 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 values. χ_1 dihedral angles, when obtained (Table I), 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 previously obtained (Kochoyan et al., 1991). Experimental restraints are provided as supplemental information (see paragraph at end of paper regarding supplementary material).

DG/SA Structure Calculations. These were performed with a new set of DG/SA programs designated DG-II (Havel, 1991). This set is based on the EMBED algorithm (Crippen, 1981; Havel et al., 1983; Havel & Crippen, 1988) and incorporates more recent improvements in sampling and convergence (Havel & Wuthrich, 1984; Pardi et al., 1988; Blaney, 1990; Havel, 1990). The maximum NOE restraint violation among the final DG/SA ensemble was 0.12 Å; the most consistent DG/SA structure exhibited an average NOE violation of 0.002 Å and a maximum NOE violation of 0.04 Å. RMS deviations are given in Figure 4.

Restrained Molecular Dynamics. The DG/SA structures obtained from DG-II were subject to restrained energy minimization and restrained molecular dynamics (RMD) in vacuo with the program XPLOR (A. Brunger, Yale University), essentially as described by Clore et al. (1985). The empirical energy function is as described for the CHARMM program (Brooks et al., 1983). This protocol has three parts: (a) initial restrained energy minimization (100 steps using the Powell algorithm), (b) 5-ps RMD trajectory (step size 1 fs) at 300 K, and (c) an additional 100 steps of restrained energy minimization as in (a). In the calculations all atoms of the protein were represented, including explicit hydrogens. NOE restraints were introduced as a square well from 0 to X Å (X = 2.7, 3.3, or 4.3 Å for strong, medium, or weak NOEs, respectively) with an upper harmonic barrier (force constant 40 kcal/Å^2). To avoid aberrant salt-bridge formation, the side chains of arginine, lysine, aspartic acid, and glutamic acid were assumed to be uncharged; for other electrostatic terms a dielectric constant of 80 was assumed. Zn–S (cysteate) and Zn–N (histidine) distances were restrained by a harmonic potential with a force constant of 200 kcal/Å^2; planarity of the Zn–histidine bond was restrained by an improper dihedral force constant of 100 kcal/rad^2. The maximum NOE restraint violation among the RMD ensemble was 0.12 Å; the most consistent RMD structure exhibited an average NOE violation of 0.003 Å and a maximum NOE violation of 0.07 Å.

Table I: Stereospecific Assignments (ppm)

residue	β_2	β_3	χ_1[a]
Tyr3	2.75	2.85	−60
Cys5	2.75	3.48	180
Tyr7	1.2	2.57	−60
Tyr10	2.73	2.05	180
Ser12	3.43	3.48	−60
Glu14	2.55	2.7	180
Leu18	1.48	2.03	180
His21	2.6	3.55	180
Ile22	1.92	1.03[b]	−60
Lys25	0.48	0.67	−60

[a] The χ_1 = −60° (t^2g^3) rotamer corresponds to a situation in which (i) $^3J_{\alpha\beta}$ coupling constants for the β and β' protons were small (<6 Hz) and large (>9.5 Hz), respectively, (ii) α to β and β' NOEs were strong and weak, respectively, and (iii) NH to β and β' NOEs were weak and strong, respectively. We then assign $\beta_2 = \beta'$ and $\beta_3 = \beta$ in the convention of Wuthrich (1986). The χ_1 = 180° rotamer (g^2t^3) corresponds to a situation in which (i) and (ii) are the same as above but in which in (iii) both NH to β and β' NOEs were strong. In this case, we assign $\beta_2 = \beta$ and $\beta_3 = \beta'$. [b] For Ile the proton is always β_2 at the C_β chiral center. The χ_1 restraint is based on the strong $^3J_{\alpha\beta}$ coupling constant (11 Hz).

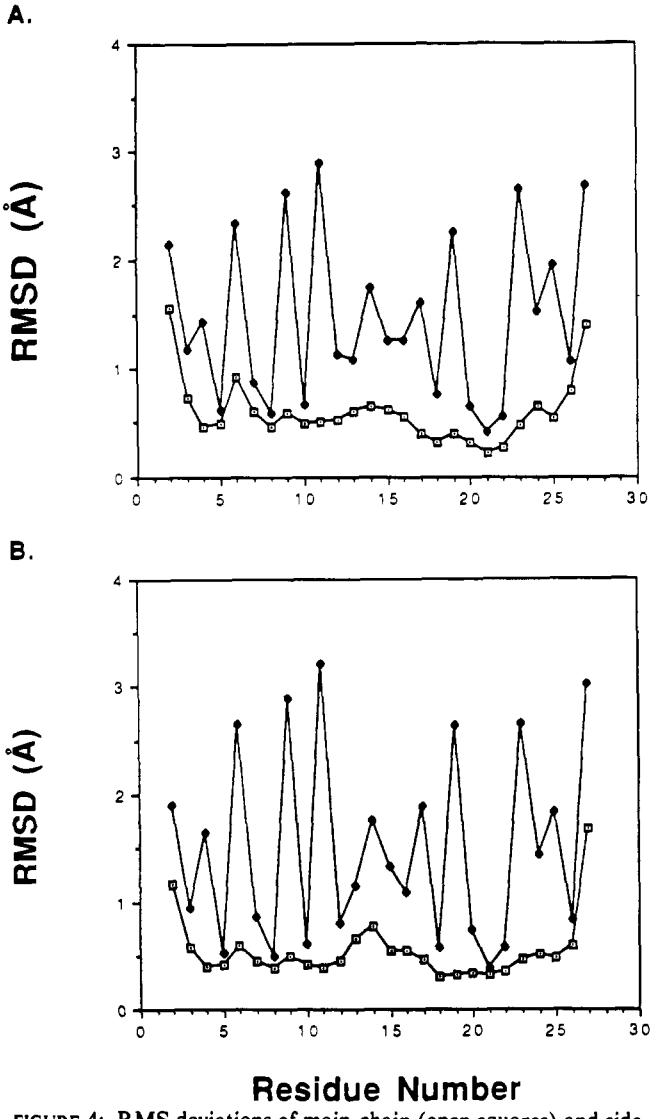

FIGURE 4: RMS deviations of main-chain (open squares) and side-chain (closed diamonds) atoms as a function of residue number. (A) RMS deviations were calculated on the basis of the average of pairwise alignments of the 13 DG/SA structures (step ii in text); pairs were aligned according to all main-chain atoms (residues 2–27). (B) Corresponding RMS deviations for the RMD ensemble.

RESULTS

(I) Synthesis and NMR Characterization of Labeled Zn Finger

Selective Deuteration of Y10. The 1D ^{1}H NMR aromatic spectrum of ZFY-6T is shown in panel A of Figure 2. Assignments have been obtained by the sequential method as described previously (Kochoyan et al., 1991). The aromatic resonances of Y10 exist in two states: a major environment (>85%), in which the ortho and meta resonances are shifted upfield and are unresolved at 6.4 ppm, and a minor environment (<15%), which appears as a broad cross-peak in TOCSY spectra (data not shown). Only the predominant conformation is considered in this and the preceding study. The composite Y10 ring resonance at 6.4 ppm exhibits extensive nonlocal NOE interactions, as shown in panels B and D of Figure 3. Since these cannot be ascribed to the ortho or meta protons individually, large distance-bound corrections (3.4 Å) were introduced in previous DG/SA calculations (supplementary material; Kochoyan et al., 1991).

Isotope-Assisted Assignment of Y10-Specific NOESY Cross-Peaks. Selective deuteration of the ortho protons of Y10 was accomplished by peptide synthesis, using an appropriately labeled precursor at cycle 20 (arrow in Figure 1; see Materials and Methods). The 1D ^{1}H NMR spectrum of the labeled Zn finger is shown in panel b of Figure 2. As expected, the composite Y10 resonance at 6.4 ppm is observed to be reduced ≈50% in amplitude; in addition, the broad minor resonance at 6.7 ppm is also attenuated, confirming its assignment to the same site. The corresponding NOESY spectra are shown in panels A and C of Figure 3. Comparison of cross-peak intensity in the spectra of the labeled and unlabeled fingers (relative to NOEs between unlabeled sites) enables Y10-related NOEs to be classified as involving the ortho protons, meta protons, or both (Table II; see below).

(II) Refinement of Solution Structure

Initial DG/SA Model. Eighteen preliminary DG/SA structures were calculated without using information obtained from isotopic labeling (Table IIA); eight did not converge and were discarded. The remaining ten are shown in panel A of Figure 5 and exhibit the classical $\beta\beta\alpha$ structural motif as originally predicted (Berg, 1988; Gibson et al., 1988). A distortion in α-helical geometry is observed in association with variant ligand spacing HX_4H. In this model 205 NOE restraints were obtained (supplementary material); of these 72 were sequential, 47 were short range (between residues 2–4 positions apart in the sequence), and 56 were long range (>4 residues in the sequence). In addition, 8 intraresidue NOEs were used in selected cases to restrain the side-chain configuration. NOE restraints involving the unresolved ring protons of Y10 included distance-bound corrections (Table IIA). J coupling constants were used to provide eight ϕ restraints, ten χ_1 restraints, and one χ_2 restraint (Table I). For DG/SA calculations the Zn^{2+}–Sγ distances were constrained to be 2.3 ± 0.1 Å and the Zn^{2+}–Nϵ distances to be 2.0 ± 0.1 Å; the zinc was assumed to lie in the plane of the imidazole rings. The bond angles of the Zn^{2+}-coordinating tetrahedron were constrained to be 109 ± 10°. 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.

Of the ten preliminary DG/SA structures, the RMS deviation among backbone atoms was 0.6 Å. However, the precision of the initial structure was limited by ambiguity among NOEs involving the aromatic ring of Y10 (side-chain RMSD 1.2 Å), as the four ring protons have identical chemical shifts (Kochoyan et al., 1991). The multiple configurations occupied by Y10 in the initial ensemble are shown in solid line in Figure 5A.

Inclusion of Additional Y10 Restraints. Comparison of the NOESY spectra of the labeled and unlabeled fingers provides additional Y10-specific restraints, as summarized in Table II. This added information was included to obtain a refined model in the three steps. (i) Ten intermediate DG/SA structure calculations were initiated by using the data in Table IIB. Of these, two did not converge and were discarded. The environments of Y10 in the remaining eight structures were analyzed in detail to further classify ortho-specific NOEs as involving the H_2 proton, the H_6 proton, or both and likewise to classify the meta-specific NOEs as involving the H_3 proton, the H_5 proton, or both. This presumptive classification is given in Table IIC. (ii) Eighteen refined DG/SA structures were calculated by using the data in Table 2C. Of these, five did not converge and were discarded. The refined ensemble exhibited an RMS deviation for backbone atoms (residues 2–27) of 0.6 Å; the RMS deviation of all atoms (residues 2–27) was 1.0 Å. RMS deviations by residues are shown in Figure 4A.

FIGURE 5: (A) Backbone configuration (dashed line) and position of Y10 (solid line) in the initial DG/SA model ZFY-6T. Ten DG/SA structures are shown. (B) Backbone configuration (dashed line) and position of Y10 (solid line) in the RMD-refined ZYF-6T model. Thirteen structures are shown. (C) Well-defined side chains (red) in the hydrophobic core in the RMD-refined model include Y3, C5, Y7, C8, Y10, L18, T20, H21, I22, and H26. The orientation of S12 on the protein surface is also well-defined (not shown).

Table II

residue	atom	residue	atom	upper bound	comments
(A) List of NOEs Involving the Unresolved Aromatic Protons of Y10 (designated ARO)					
Cys5	HN	Tyr10	ARO	7.7	
Tyr10	ARO	His21	HE2	7.7	
Tyr10	ARO	His21	HD4	7.7	
Tyr10	ARO	Arg11	HN	7.7	
Tyr10	ARO	Arg11	HN	7.7	
Tyr10	ARO	Arg11	HA	7.7	
Tyr10	ARO	Leu18	HA	7.7	
Tyr10	ARO	Leu18	MD1	7.8	M stereospecific
Tyr10	ARO	Leu18	MD1	7.8	M stereospecific
Tyr10	ARO	Leu18	HB1	7.7	stereospecific
Tyr10	ARO	His21	HB1	7.7	stereospecific
Tyr10	ARO	His21	HB2	7.7	stereospecific
Tyr10	ARO	Ser12	HB1	7.7	
Tyr10	ARO	Ser12	HB2	7.7	
(B) List of NOEs Involving the Aromatic Ring of Y10 Obtained after the Selective Deuteration Experiment					
Cys5	HN	Tyr10	CG1	6.4	
Tyr10	CG1	His21	HE2	6.4	
Tyr10	CG1	His21	HD4	6.4	
Tyr10	CG1	Arg11	HN	7.7	*a*
Tyr10	CZ4	Arg11	HN	7.7	*a*
Tyr10	CG1	Arg11	HA	6.4	
Tyr10	CZ4	Leu18	HA	6.4	
Tyr10	CG1	Leu18	MD1	6.5	M stereospecific
Tyr10	CZ4	Leu18	MD1	6.5	M stereospecific
Tyr10	CZ4	Leu18	HB1	6.4	stereospecific
Tyr10	CZ4	His21	HB1	6.4	stereospecific
Tyr10	CZ4	His21	HB1	6.4	stereospecific
Tyr10	CZ4	His21	HB2	6.4	stereospecific
Tyr10	CZ4	Ser12	HB1	6.4	
Tyr10	CZ4	Ser12	HB2	6.4	
(C) List of NOEs Involving the Aromatic Ring of Y10 Used for Refined DG/SA and RMD Structure Calculations					
Cys5	HN	Tyr10	CG1	6.4	*b*
Tyr10	CG1	His21	HE2	6.4	*b*
Tyr10	HD2	His21	HD4	4.3	
Tyr10	ARO	Arg11	HN	7.7	*a*
Tyr10	CG1	Arg11	HA	6.4	*b*
Tyr10	HE3	Leu18	HA	4.3	
Tyr10	HD2	Leu18	MD1	4.4	M stereospecific
Tyr10	HE3	Leu18	MD1	4.4	M stereospecific
Tyr10	He3	Leu18	HB1	4.3	stereospecific
Tyr10	HE3	His21	HB1	4.3	stereospecific
Tyr10	HE3	His21	HB2	4.3	stereospecific
Tyr10	HE3	Ser12	HB1	4.3	
Tyr10	HE3	Ser12	HB2	4.3	

[a] Since this very weak NOE was not observed in the deuterated sample (due to lower intrinsic signal to noise), the distance-bound correction of 3.4 Å is maintained. [b] The geometry of this packing interaction in the intermediate model (step ii under Results) does not allow a presumptive side-specific assignment to be established. As a correction for the observed degeneracy of the two ortho protons (and, likewise, of the two meta protons) of tyrosine, distance restraints involving the ortho protons are referred to the CG1 position (and those involving the meta protons are referred to the CZ4 position) with addition of a 2.1-Å distance-bound correction as described (Wuthrich, 1986). M: methyl correction of 1 Å was applied (see Materials and Methods).

The backbone configuration is the same as the initial model; however, the position of the Y10 side chain and its interactions with neighboring residues are significantly better defined (Y10 side chain RMSD 0.6 Å). (iii) RMD calculations may, in principle, be used to further explore configuration space in the neighborhood of the DG/SA structures. Such refinment is designed to relieve unfavorable chemical features (e.g., non-ideal bond lengths, bond angles, distortions of planarity, and van der Waals overlap) and to include the stabilizing effects of possible hydrogen bonding. Each of the DG/SA structures obtained in (ii) was subject to an RMD protocol with the program XPLOR (A. Brunger, Yale University) as described (Clore et al., 1985; see Materials and Methods). Only local changes were introduced by the RMD/energy minimization procedure; the average displacement between input and output structures was <0.5 Å. Although the DG/SA and RMD ensembles exhibit equivalent consistency with the experimental restraints, the latter shows significant improvement in local chemical "reasonableness"; individual contributions to the empirical energy before and after RMD refinement are given in Table III. RMS deviations of the RMD ensemble are similar to those of the input DG/SA ensemble (Figure 4B). Following RMD, the orientations of internal side chains are somewhat better defined, presumably as a result of the inclusion of dihedral and hydrogen-bonding terms in the empirical energy function; the orientations of external side chains are less well-defined, suggesting more extended sampling properties in the absence of restraints. The configuration of Y10, shown in Figure 5B, remains well-defined (side-chain RMSD 0.6 Å), as obtained in (ii).

Table III: Average Calculated Empirical Energy of ZFY-6T following DG/SA and RMD[a]

energy term	DG/SA	RMD	Δ
NOE residual violation[b]	1.6	1.7	−0.1
dihedral restraint violation[b]	0.0	0.0	−0.0
bonds	47.6	7.1	−40.5
hydrogen bonds	−8.6	−41.5	−32.9
angles	136.0	90.4	−45.6
dihedral	180.5	89.5	−90.5
improper dihedral	38.9	2.9	−36.0
van der Waals	−104.5	−153.0	−48.5
electrostatic[c]	4.2	2.2	−2.0
total	288.0	−1.2	−289.2

[a] The results are given in kilocalories per mole; the DG/SA and RMD ensembles correspond to steps ii and iii, respectively, in part II under Results. [b] Both the DG/SA and RMD ensembles satisfied the input restraints (supplementary material table) to an equal extent. [c] The electrostatic term is relatively small because the charged amino acids (K, R, E, and D) are assumed to be uncharged and because a dielectric constant of 80 is otherwise employed.

Backbone configurations in the ensemble of RMD-refined structures (Figure 5B) are similar to those in the initial DG ensemble (Figure 5A). The configuration of the "fingertip" (residues 12–14; i.e., the loop between the N-terminal β-sheet and C-terminal α-helix) is somewhat better defined. The following side chains are well oriented and form a compact hydrophobic core (Figures 4 and 5C): Y3, C5, Y7, C8, Y10, L18, T20, H21, I22, and H26 (S12 is also well-defined near the protein surface). Fifteen hydrogen bonds are identified following RMD and restrained minimization; these are specified in Table IV by donor, acceptor, length, angle, and empirical contribution to the calculated electrostatic energy. Presumptive hydrogen bonds may be classified as β-sheet related, α-helix related, and other. Three features are noteworthy: (i) The β-sheet geometry satisfies a conventional hydrogen-binding scheme, a point at issue among other Zn fingers (Lee et al., 1989; Klevit et al., 1990; Omichinski et al., 1990). (ii) The distortion in α-helical geometry associated with the HX_4H ligand spacing (see above) makes possible two sets of bifurcating hydrogen bonds (between the amide proton of T24 and the carbonyl oxygens of T20 and H21 and between the carbonyl of H21 and the amide protons of T24 and K25). This hydrogen-bonding scheme has not previously been observed among HX_3H or HX_5H metal-binding sites. (iii) The *p*-hydroxyl group of Y10 participates in hydrogen bonds with the side chain of Ser12 and in some structures also with the side chain of Asn17. The latter has been identified as a specific

Table IV: Hydrogen Bonds Predicted by RMD Refinement[a]

	donor		acceptor		no. of structures involved	energy (kcal/mol)	distance (Å)	angle [Δ (deg)][b]
3	N	HN	O=C	12	12	−2.6	3	9
4	NE2	HE	O=C	5	8	−1.2	3.1	14
5	N	HN	O=C	10	12	−2.0	3	26
10	OH	HH	OG	12	11	−2.4	3.2	25
12	N	HN	O=C	3	13	−2.6	3	9
12	OG	HG	OH	10	8	−2.6	3.2	21
17	ND	HD	OH	10	4	−1.2	3.1	40
19	N	HN	O=C	15	13	−2.8	3	10
20	N	HN	O=C	16	13	−2.8	3.1	5
21	N	HN	O=C	17	12	−2.1	3	9
22	N	HN	O=C	18	13	−3.0	3	10
23	N	HN	O=C	19	13	−2.7	3	11
24	N	HN	O=C	20	9	−1.9	2.9	28
24	N	HN	O=C	21	11	−0.1	3.1	40
25	N	HN	O=C	21	13	−1.9	2.8	25

[a] Presumptive hydrogen bonds are listed only if present in more than 4 out of 13 structures. Carbonyl oxygen acceptors are designated O=C. [b] Δ (deg) is defined as deviation from linearity (in degrees).

DNA contact residue in other systems (Nardelli et al., 1991; Pavletich & Pabo, 1991).

It is important to consider whether the empirical energy function used in the RMD protocol may introduce apparent hydrogen bonds not directly implied by the experimental NMR data. To address this issue, the geometry of the potential hydrogen-bond donors and acceptors was also evaluated in the DG ensemble following step ii (above). We find that each of the hydrogen bonds involving the polypeptide backbone—including the bifurcated hydrogen-bonding scheme associated with HX_4H ligand spacing—is compatible with each structure in the DG ensemble; a table of maximum and average hydrogen-bond lengths, angles, and empirical energy is provided as supplementary material. The RMD protocol in essence provided a local improvement in the ideality of these hydrogen bonds within the limits of the experimental restraints. Hydrogen bonds involving side chains, variably present in the RMD ensemble, are also only variably present in the DG ensemble. The latter interactions should therefore be considered as stereochemically permitted but hypothetical.

Comparison to Other Fingers. The backbone configuration of ZFY-6T (shown in green in Figure 6A) is similar to that of a representative Xfin-31 structure (shown in blue; Lee et al., 1989) and of a Zn finger from a human enhancer binding protein (shown in red; Omichinski et al., 1990). Local differences are observed in the details of the β-sheet and in association with variant HX_3H, HX_4H, or HX_5H ligand spacing.

Each of these fingers exhibits a well-defined hydrophobic core involving analogous residues. However, the even-specific residue Y10 differs from the consensus aromatic residue F12 (observed in the earlier structures) in both internal orientation and surface topography. The internal orientations of Y10 and F12 in representative structures are shown in Figure 7 relative to conserved "framework" residues L18 and H21. Whereas F12 packs primarily against L18, Y10 is nearly stacked against the H21 imidazolic ring and approaches L18 in a different orientation. The importance of possible side-chain hydrogen bonds (Y10–S12 and/or Y10–N17; see above) in stabilizing this structure is unclear and may in the future be investigated by comparative studies of finger analogues. Interestingly, the solvent-exposed surface of Y10 (shown in red in Figure 6B) is contiguous with a surface defined by residues implicated in other systems (Nardelli et al., 1991; Pavletich & Pabo, 1991) in specific DNA recognition (residues 14, 17, 20, and 21; shown in blue in the figure). In striking contrast, residue 10 in other Zn fingers (Lee et al., 1989; Omichinski et al., 1990) adopts a different orientation and so contributes to a noncontiguous region of the protein surface (Figure 6C,D; color code as in panel B). The consensus residue R10 in classical Zn fingers has been shown to contact a phosphate in a protein–DNA complex (Pavletich & Pabo, 1991). Because the sequence of odd-numbered ZFY domains is similar to that of Xfin-31 and other consensus Zn fingers (Page et al., 1987), these results suggest that the strict conservation of Y10 among domains 4, 6, and 12 and alternation between Y10- and F12-containing domains in the ZFY two-finger repeat (DiLella et al., 1990) will have implications for the mechanism of DNA recognition. Possible models of the ZFY–DNA complex have previously been discussed (Kochoyan et al., 1991).

Discussion

The Zn finger is a metal-binding site containing two cysteine and two histidine residues with characteristic spacing ($CX_{2,4}CX_{12}HX_{3-5}H$; Brown et al., 1985; Miller et al., 1985). Peptide models of the Zn finger provide independent folding units (Frankel et al., 1987), whose structures may be determined by 2D NMR spectroscopy (Parraga et al., 1988; Lee et al., 1989; Klevit et al., 1989; Omichinski et al., 1990; Kochoyan et al., 1991). These studies reveal a compact globular domain in which the divalent metal is encaged; the N-terminal portion (containing the conserved cysteines) forms a β-sheet and turn, and the C-terminal portion (containing the conserved histidines) forms an α-helix as previously proposed (Berg, 1988; Gibson et al., 1988). Other types of metal-binding motifs have been identified, including retroviral and steroid receptor related fingers, which exhibit distinct three-dimensional structures (Summers et al., 1990; Hard et al., 1990; Schwabe et al., 1990).

A majority of classical Zn-finger sequences exhibit a single-finger repeat (Gibson et al., 1988), consistent with a proposed model of DNA recognition (model I; Fairell et al., 1986). This model, in which successive Zn fingers are proposed to "wrap around" the major groove, has been refined by Berg (1990) on the basis of the 2D NMR structure of the isolated Zn finger (Lee et al., 1989; Klevit et al., 1990); its major features have recently been confirmed by crystallographic studies of a specific three-finger–DNA complex (Pavletich & Pabo, 1991). This mode of DNA binding may not apply, however, to families of Zn-finger proteins that exhibit second-order repeats, as defined by systematic difference between odd- and even-numbered domains. Examples include an extensive collection of Kruppel-related sequences in *Xenopus laevis* (Nietfeld et al., 1989) and the ZFY-related gene family (Page et al., 1987; Schneider-Gadicke et al., 1989). Odd-numbered ZFY domains 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  20   25
```

whereas the even-numbered fingers follow the variant template:

```
(N) K-ArQC-Y/HC-Ar-S/T----EL--H----HS/TK
    1   5      10     15    20   25
```

where h denotes a hydrophobic residue and Ar denotes an aromatic group (His, Phe, or Tyr). The even template differs from that of the odd in the position of the conserved central aromatic group (Ar10 or F12; underlined above), in the spacing between histidines, and in various differences in amino

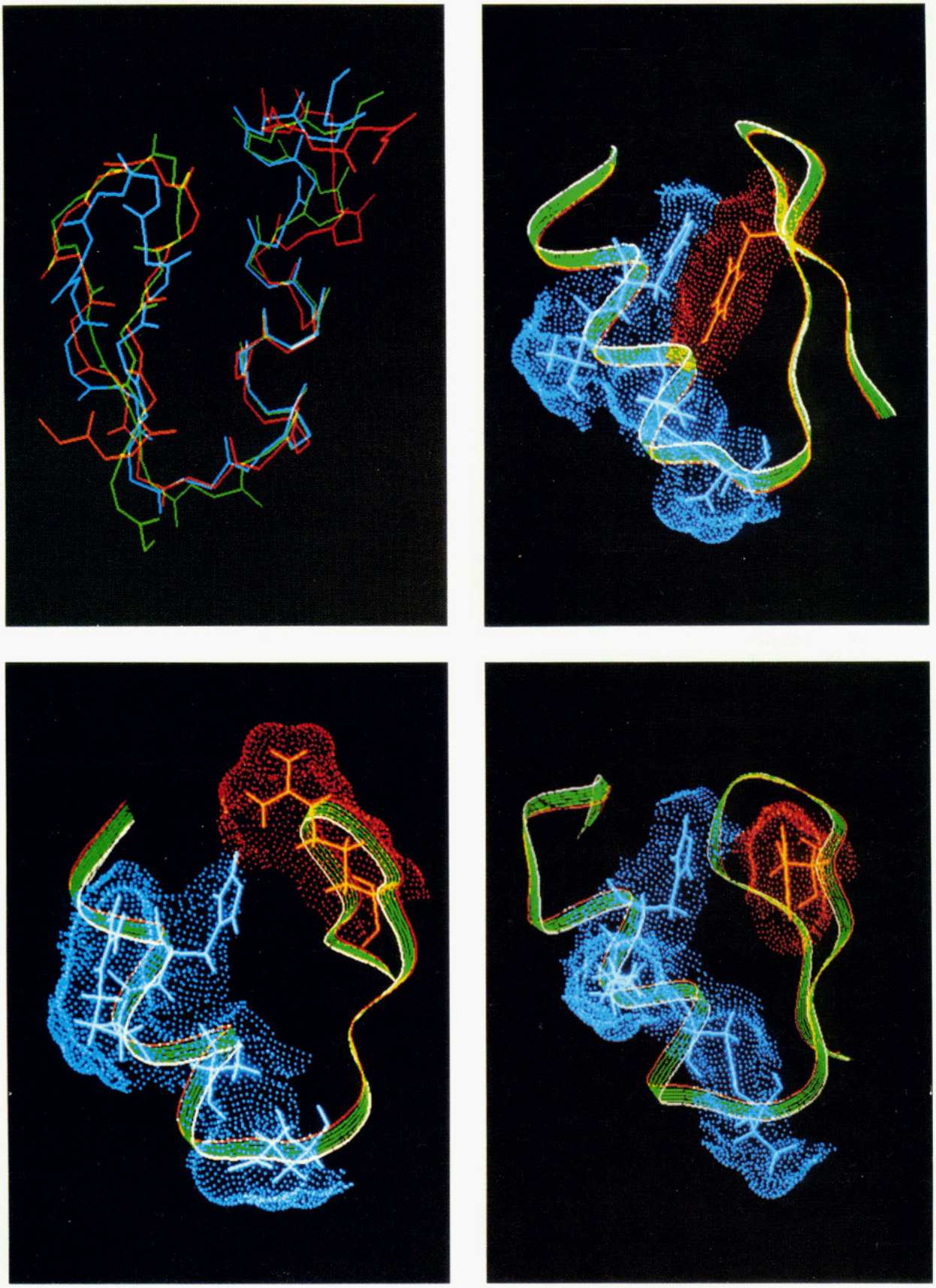

FIGURE 6: (A, top left) Alignment of ZFY-6T (green), Xfin-31 (blue; Lee et al., 1989), and a finger derived from a human enhancer binding protein (red; Omichinski et al., 1990) showing similar elements of secondary structure ($\beta\beta\alpha$ motif). The alignment was based on a least-squares fit of the main-chain atoms of α-helical residues 17–21 (numbering scheme as in Figure 1). (B, top right) Surface and ribbon representations of ZFY-6T showing putative specificity-determining residues (blue; Pavletich & Pabo, 1991) and the contiguous contribution of Y10 (red). (C, Bottom left, and D, bottom right) Corresponding ribbon and dot-surface representations of Xfin-31 (Lee et al., 1989) and the enhancer-binding finger (Omichinski et al., 1990), respectively. In both structures residue 10 (red) is oriented outward and is not contiguous with the putative specificity-determining surface (blue).

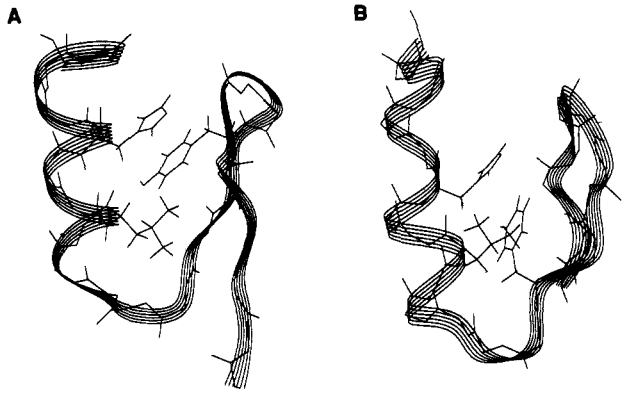

FIGURE 7: Relative orientation of central aromatic residues Y10 (panel A) and F12 (panel B) in ZFY-6T and Xfin-31, respectively, as aligned according to their backbone elements (ribbon). Conserved "framework" residues L18 and H21 are also shown.

acid preferences at other sites (Page et al., 1987; DiLella et al., 1990). ZFY-6T, the object of the present study, provides a representative example of an even (nonconsensus) domain in the ZFY two-finger repeat (Page et al., 1987; Weiss et al., 1990).

Patterns of conservation and variation among Zn-finger sequences are of general interest as models of templates for protein folding. Although it is likely that individual Zn fingers differ in detail, accurate and precise knowledge of their structures and interactions will be required to understand the informational content of these templates. There presently is disagreement in the literature regarding the extent of flexibility in the hydrophobic core of classical Zn fingers: three independent structures provide different views (Lee et al., 1989; Klevit et al., 1990; Omichinski et al., 1990). Although the interior structures of two of these domains are well-defined (Lee et al., 1989; Omichinski et al., 1990), the central aromatic group of ADR1b (F12 in the present number scheme; Figure 1) was described as projecting in different directions in the three DG structures shown (Klevit et al., 1990). In the X-ray structure of a Zn-finger protein–DNA complex, the homologous residues are also observed to occupy well-defined orientations (Pavletich & Pabo, 1991).

In our initial DG/SA structure of ZFY-6T (Kochoyan et al., 1991), imprecision was also observed in the configuration of Y10; however, this appeared likely to be due to resonance overlap rather than to actual disorder in the peptide. To address this issue, we have used selective deuterium labeling to assign previously unresolved NOEs. These additional restraints permitted a refined solution structure to be obtained. Previously ambiguous NOEs involve a conserved aromatic ring (Y10) whose interactions in the hydrophobic core contribute to the stability of the domain (Weiss & Keutmann, 1990). In the refined structure the hydrophobic core is well-ordered; in particular, the configuration of Y10 is well-defined. Its contacts with conserved residues L18 and H21 differ in detail from that of a conserved aromatic residue at an alternative site (F12) in consensus Zn fingers (Lee et al., 1989; Pavletich & Pabo, 1991). However, the latter sequences differ from ZFY-6T at multiple positions, complicating the interpretation of local differences in structure. Such comparisons may more meaningfully be made by analysis of related analogues, such as a designed "aromatic swap" revertant (Weiss & Keutmann, 1990). Such studies are in progress and will be described elsewhere. The present results suggest that alternation between even- and odd-sequence patterns alters the topology of the putative DNA-binding surface, implying that the ZFY two-finger repeat will have implications for the mechanism of DNA recognition by this class of Zn-finger proteins.

ACKNOWLEDGMENTS

We thank K. A. Mason for technical assistance, C. A. Dahl for advice regarding peptide synthesis, T. F. Havel for the program DG-II and A. T. Brunger for XPLOR, D. C. Page and P. A. Sharp for helpful discussion and communication of results prior to publication, G. M. Clore and A. M. Gronenborn for the coordinates of an enhancer-binding Zn finger, N. Pavletich and C. O. Pabo for the coordinates of the Zif268–DNA complex, and D. Case and P. E. Wright for the coordinates of Xfin-31 and helpful discussion.

SUPPLEMENTARY MATERIAL AVAILABLE

A list of metal-related and NOE restraints and a table of candidate hydrogen bonds in the DG ensemble, giving average and maximum hydrogen-bond lengths and angles (8 pages). Ordering information is given on any current masthead page.

REFERENCES

Barany, G., & Merrifield, R. B. (1979) in *The Peptides* (Gross, E., & Meienhofer, J., Eds.) Vol. 2, pp 1–284, Academic Press, New York, NY.

Berg, J. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 99–487.

Blaney, J. (1990) *The DGEOM Program*, Quantum Chemistry Program Exchange, Department of Chemistry, University of Indiana, Bloomington, IN.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. O., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

Brown, R. S., Sanders, C., & Argos, S. (1985) *FEBS Lett. 186*, 271–274.

Churchill, M. E. A., Tullius, D. T., & Klug A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5528–5532.

Clore, G. M., Gronenborn, A. M., Brunger, A., & Karplus, M. (1985) *J. Mol. Biol. 186*, 435–455.

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., 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.

Hard, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., & Kaptein, R. (1990) *Science 249*, 157–160.

Havel, T. F. (1990) *Biopolymers 29*, 1565–1585.

Havel, T. F. (1991) in *Progress in Biophysics and Molecular Biology* (Noble, D., & Blundell, T. E., Eds.) Pergamon 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.

Karplus, M. (1959) *J. Phys. Chem. 30*, 11–15.

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., Havel, T. F., Nguyen, D., Dahl, C. E., Keutmann, H. T., & Weiss, M. A. (1991) *Biochemistry 30*, 3371–3386.
Koopman, P., Gubbay, J., Colignon, J., & Lovell-Badge, R. (1989) *Nature 342*, 940–942.
Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., & Wright, P. E. (1989) *Science 245*, 635–637.
Miller, J., MacLachlan, A. D., & Klug, A. (1985) *EMBO J. 4*, 1609–1614.
Nardelli, J. Gibson, T. J., Vesque, C., & Charnay, P. (1991) *Nature 349*, 175–178.
Nietfeld, W., El-Baradi, T., Mentzel, H., Pieler, T., Koster, M., Poting, A., & Knochel, W. (1989) *J. Mol. Biol. 208*, 639–659.
Omichinski, J. G., Clore, G. M., Apella, E., Sakaguchi, K., & Gronenborn, A. M. (1990) *Biochemistry 29*, 9324–9334.
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.
Pavletich, N. P., & Pabo, C. O. (1991) *Science 252*, 809–817.
Schneider-Gadicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R., & Page, D. C. (1989) *Cell 57*, 1247–1258.
Schwabe, J. W. R., Neuhaus, D., & Rhodes, D. (1990) *Nature 348*, 458–461.
Stewart, J. M., & Young, J. D. (1984) *Solid-Phase Peptide Synthesis*, Raven Press, New York, NY.
Summers, M. F., South, T. L., Kim, B., & Hare, D. R. (1990) *Biochemistry 29*, 329–338.
Tregear, G. W., van Reitschoten, J., Sauer, R. T., Niall, H. D., Keutmann, H. T., & Potts, J. T. (1977) *Biochemistry 16*, 2817–2823.
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.
Wuthrich, K. (1989) *Methods Enzymol. 177*, 125–131.

Nucleotide Positions Responsible for the Processivity of the Reaction of Exonuclease I with Oligodeoxyribonucleotides†

Richard S. Brody
Department of Pharmacology and Toxicology, Battelle, Columbus, Ohio 43201
Received February 8, 1991; Revised Manuscript Received May 6, 1991

ABSTRACT: The processive hydrolysis of single-stranded oligodeoxyribonucleotides by exonuclease I from *Escherichia coli* has been investigated. Oligodeoxyribonucleotides and their analogues, which contain either an abasic site or a methylphosphonate internucleotide linkage, were partially hydrolyzed by exonuclease I. The relative dissociation constant for the enzyme and each oligomeric product was calculated from the concentration of that oligomer found in solution and hence released by the enzyme before complete hydrolysis. The results have led to a characterization of the two oligodeoxyribonucleotide domains that bind to exonuclease I. The first domain, which begins at the reactive 3′-terminal phosphodiester and extends to the 7th nucleoside base, requires both phosphodiester monoanions and base residues for its interaction with the enzyme. The second domain includes phosphodiester monoanions in positions 9–13 from the 3′-terminus but does not require nucleoside bases. Methylphosphonate substitutions indicate that only two or three of these phosphodiesters, in variable positions, must remain anionic in order to obtain full enzyme binding. The residues between the two binding domains do not play a significant role in the enzyme–oligomer interaction.

Exonuclease I from *Escherichia coli* catalyzes the hydrolysis of mononucleotide 5′-phosphates from the 3′-terminus of single-stranded DNA (Lehman, 1960; Lehman & Nussbaum, 1964) in a highly processive reaction (Thomas & Olivera, 1978; Brody et al., 1986). Experiments with phosphorothioate analogues have shown that the stereochemical course of the reaction is inversion of configuration at phosphorus and that phosphodiester bond cleavage is at least partially rate limiting (Brody & Doherty, 1985). Kinetic investigations indicate that the maximum rate of hydrolysis is independent of polymer size, and competition experiments have shown that the enzyme associates directly with the 3′-terminus of a polydeoxyribonucleotide (Brody et al., 1986).

An important feature of many enzymes that react with nucleic acids is their ability to remain bound to the polymeric substrate between successive catalytic events (Kornberg, 1980). Exonucleases that react by such a processive mechanism have

†Supported by Grant GM-33266 from the National Institute of General Medical Sciences.

0006-2960/91/0430-7072$02.50/0