Novel Zinc Finger Motif in the Basal Transcriptional Machinery: Three-Dimensional NMR Studies of the Nucleic Acid Binding Domain of Transcriptional Elongation Factor TFIIS†,‡

Xiuqi Qian,§,# Shai N. Gozani,§ HoSup Yoon,∥ ChoonJu Jeon,∥ Kan Agarwal,∥,⊥ and Michael A. Weiss*,§,○

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, and Departments of Biochemistry and Molecular Biology and of Chemistry, The University of Chicago, Chicago, Illinois 60637

Received February 22, 1993; Revised Manuscript Received April 20, 1993

ABSTRACT: Transcriptional elongation provides a key control point in the regulation of eukaryotic gene expression. Here we describe homonuclear and ^{15}N-heteronuclear 3D NMR studies of the nucleic acid binding domain of human transcriptional elongation factor TFIIS. This domain contains a Cys_4 Zn^{2+}-binding site with no homology to previously characterized Cys_4, Cys_6, or Cys_2-His_2 Zn fingers. Complete ^1H and ^{15}N NMR resonance assignment of a 50-residue TFIIS peptide–Zn^{2+} complex is obtained. Its solution structure, as determined by distance geometry/simulated annealing (DG/SA) calculations, exhibits a novel three-stranded antiparallel β-sheet (designated the Zn ribbon). Analogous sequence motifs occur in a wide class of proteins involved in RNA or DNA transactions, including human basal transcriptional initiation factor TFIIE. A three-dimensional model of the TFIIE Cys_4 domain is obtained by DG-based homology modeling. The role of the TFIIS Zn ribbon in the control of eukaryotic transcriptional elongation is discussed.

Gene expression is regulated at several transcriptional and translational control points. Recent interest has focused on the role of sequence-specific DNA-binding proteins (specific transcription factors) in eukaryotic transcriptional initiation. These proteins define families of related structures (Pabo & Sauer, 1992) and interact with a promoter-engaged assembly of RNA polymerase and accessory initiation proteins (basal transcription factors). Complementary mechanisms to regulate transcriptional elongation are well-described in prokaryotic operons (Yanofsky, 1988). Because central features of such models (e.g., polycistronic operons and coupled transcription–translation) are specific to prokaryotes, analogous mechanisms of transcriptional pausing and elongation in eukaryotes are not well-understood and pose an outstanding problem (Bentley & Groudine, 1986, 1988). Such mechanisms are also of interest as potential targets for anti-neoplastic (Wright & Bishop, 1989) and anti-viral drug design (Hay et al., 1982; Hay & Aloni, 1984; Kao et al., 1987).

Transcriptional pause sites in eukaryotic genes contain specific sequence elements (Baek et al., 1986; Chinksy et al., 1989; Resnekov & Aloni, 1989) whose use may be regulated in response to signaling pathways (Mather et al., 1984; Bentley & Groudine, 1988). In the absence of an induction signal, for example, transcriptionally engaged RNA polymerase II can be blocked just downstream of a heat-shock promoter with nascent RNA stably attached (Rougvie & Lis, 1988). Although the proteins involved in such regulation *in vivo* are not known, a highly conserved elongation factor (designated TFIIS[1] or S-II) has been purified from several sources (Natori et al., 1973; Sawadogo et al., 1980; Natori, 1982; Horikoshi et al., 1984; Rappaport et al., 1987; Reinberg & Roeder, 1987; Sluder et al., 1989). TFIIS stimulates RNAP-II in the elongation step *in vitro*, permitting read-through at specific pause or termination sites (Reines et al., 1989; Siva-Raman et al., 1990; Agarwal et al., 1991). Although the molecular basis of this activity is not understood, TFIIS has recently been shown to stimulate transcript shortening by template-engaged RNA polymerase II. This reaction is Mg^{2+}-dependent and may provide an "antipausing" 3′→5′ ribonuclease activity in the elongation complex (Reines & Mote, 1993; Izban & Luse, 1992). An analogous ribonuclease activity has been shown to be stimulated by bacterial elongation factor GreA (Borukhov et al., 1992).

The sequence of TFIIS cDNA has been determined from mouse (Hirashima et al., 1988; Kanai et al., 1991), man (Yoon

† This work was supported by grants from the National Institutes of Health (K.A.), the Elsa U. Pardee Foundation, and the Council for Tobacco Research (M.A.W.). M.A.W. is an Established Investigator of the American Heart Association. This paper is dedicated to the memory of Dr. Leo J. Neuringer, founding director of the Biological NMR Resource at the MIT-Francis Bitter National Magnet Laboratory.

‡ Protein coordinates have been submitted to the Brookhaven Protein Data Bank under file name 1TFI.

* Address correspondence to this author at Harvard Medical School.

§ Harvard Medical School.

Present address: Fuji Immunopharmaceuticals Corp., 125 Hartwell Ave., Lexington, MA 02173.

∥ Department of Biochemistry and Molecular Biology, The University of Chicago.

⊥ Department of Chemistry, The University of Chicago.

○ Massachusetts General Hospital.

[1] Abbreviations: Ar, aromatic residue; CD, circular dichroism; DG, distance geometry; DQF-COSY, double-quantum filtered correlated spectroscopy; EDTA, ethylenediaminetetraacetic acid; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; NOESY-HMQC, three-dimensional ^1H nuclear Overhauser enhancement ^{15}N–^1H heteronuclear multiple-quantum coherence spectrum; RMS; root mean square; RMSD, RMS difference; eukaryotic RNAP-II, RNA polymerase II; ROESY, rotating frame Overhauser spectroscopy; rp-HPLC, reverse-phase high-performance liquid chromatography; rpm, revolutions per minute; SA, simulated annealing; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; TFIIE, basal transcription factor E for RNAP-II; TFIIS, transcriptional elongation factor SII for RNAP-II; TOCSY, total correlated spectroscopy; TOCSY-NOESY, three-dimensional homonuclear spectrum containing TOCSY transfer in t_1 and NOE transfer in t_2; UV, ultraviolet; 2D NMR, two-dimensional NMR; EXAFS, extended X-ray absorption fine structure; 3D NMR, three-dimensional NMR.

0006-2960/93/0432-9944$04.00/0

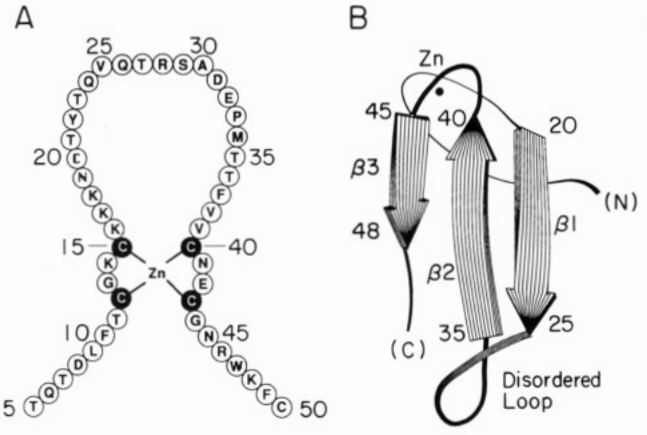

FIGURE 1: (A) Primary structure of TFIIS nucleic acid binding domain (peptide residues 5–50; protein residues 235–280). Invariant cysteines involved in tetrahedral Zn^{2+} coordination are shown as filled circles. (B) Ribbon model of TFIIS nucleic acid binding domain based on DG/RMD calculations (Qian et al., 1993). The position of Zn^{2+} and end points of β-strands 1–3 are indicated.

et al., 1991), *Drosophila melanogaster* (Marshall et al., 199), *Saccharomyces cerevisiae* (Nakanishi et al., 1992), and vaccinia virus (Ahn et al., 1991). The sequences are highly homologous. Human TFIIS contains 280 residues and *in vivo* is phosphorylated (Yoon et al., 1991). Partial proteolysis and analysis of deletion mutants demonstrate three discrete regions: a variable N-terminal domain (containing phosphorylation sites), a central domain (which binds RNAP-II), and a C-terminal domain (which binds nucleic acids; Agarwal et al., 1991). The C-terminal domain contains a novel Cys_4 motif in which a single Zn^{2+} is coordinated with tetrahedral geometry (Figure 1A; Agarwal et al., 1991).[2] Site-directed mutagenesis has demonstrated that the four conserved cysteines are involved in metal binding (C. J. Jeon, H.-S. Yoon, and K. Agarwal, manuscript in preparation).

In this paper we describe the three-dimensional solution structure of the C-terminal nucleic acid binding domain. Complete ^1H and ^{15}N NMR resonance assignment of a 50-residue fragment is obtained by homonuclear TOCSY-NOESY (Oschkinat et al., 1989a,b) and ^{15}N-edited NOESY 3D NMR methods (Clore & Gronenborn, 1991). Distance geometry/simulating annealing (DG/SA) calculations reveal a novel three-stranded antiparallel β-sheet (designated the *Zn ribbon*; Figure 1B). An analogous Cys_4 motif is predicted in eukaryotic basal transcriptional initiation factor TFIIE (Peterson et al., 1991; Ohkuma et al., 1991), subunits of RNA polymerase II, and other proteins involved in DNA or RNA transactions. A three-dimensional structure of the TFIIE domain is proposed based on homology modeling (Havel, 1990). Implications of the structure of the TFIIS nucleic acid binding domain for models of elongational control (Kassavetis & Geiduschek, 1993) are discussed. A qualitative description of the TFIIS structure will be presented elsewhere (Qian et al., 1993).

MATERIALS AND METHODS

Buffers. Buffer A (bacterial lysis buffer SW) consists of 20 mM sodium phosphate (pH 7.8), 500 mM NaCl, 10 mM β-mercaptoethanol, 5 mM benzamidine, 1 mM *N*-tosyl-L-phenylalanine chloromethyl ketone (TPCK), 1 mM N^{α}-*p*-tosyl-L-lysine chloromethyl ketone (TLCK), and 0.2 mM phenylmethanesulfonyl fluoride (PMSF). Buffer B consists of 50 mM sodium phosphate (pH 7.4), 100 mM $(NH_4)_2SO_4$, 0.2 mM Na·EDTA, 50 mM NaCl, 20 mM β-mercaptoethanol, 10 μM $ZnSO_4$, 1 mM TPCK, 1 mM TLCK, 5 mM benzamidine, and 0.2 mM PMSF. Buffer C consists of 25 mM Tris-HCl (pH 7.4), 50 mM NaCl, 0.1 mM EDTA, 10 mM β-mercaptoethanol, and 10% glycerol. Buffer D consists of 20 mM sodium phosphate (pH 6.0), 500 mM NaCl, and 10 mM β-mercaptoethanol. CD buffers consist of 50 mM Tris-HCl (pH 7.4 at 25 °C) and 0.1% acetic acid (pH 3). NMR buffer consists of 50 mM deuterated Tris-HCl (pH or p*D* 5.8 at 25 °C, direct meter reading) and 10 mM deuterated DTT. Deuterated solvents and chemicals were purchased from Cambridge Isotopes, Inc. (Woburn, MA), and Merck Isotopes, Inc. (St. Louis, MO). Buffers were deoxygenated with N_2 prior to use.

Peptide Expression. Two T7-based expression systems were constructed in *E. coli*. (i) Residues 231–280 of human TFIIS (Yoo et al., 1991) were subcloned into T7 expression vector. This construction yielded a product (designated fragment A) consisting of residues 231–280 preceded by N-terminal nonnative residues Met-Ala-Arg-Ala-Pro. The level of expression was approximately 5% as judged by SDS–PAGE. (ii) A similar region of TFIIS (residues 229–280) was recloned into a His_6-tagged T7 system to be described elsewhere. This construction yielded a product (designated fragment B) consisting of residues 229–280 preceded by N-terminal nonnative residues Met-Arg-His-His-His-His-His-His-Pro. The efficiency of expression was similar to that of the first construction. Fidelity of each plasmid construction was verified by DNA sequencing. The amino acid sequences of fragments A and B are otherwise as given in Table I. A schematic representation of the Zn-binding site (peptide residues 5–50; TFIIS residues 235–380) is shown in Figure 1A. One-dimensional ^1H NMR spectra of the two fragments are similar (supplementary material). Native and His_6-tagged peptide–Zn^{2+} complexes exhibit identical nucleic acid binding properties in gel retardation experiments (Qian et al., 1993) under conditions described (Agarwal et al., 1991).

Peptide Purification. Overexpression strains of *E. coli* were grown in Luria broth (LB) for isolation of unlabeled protein and in minimal medium (glucose, [^{15}N]-ammonium chloride, thiamine, and M9 salts) for isolation of uniformly ^{15}N-labeled peptides. Overall yield from either expression system was 5 mg/L (rich medium) and 2 mg/L (minimal medium). In each case cells were grown to an optical density (OD_{590}) of 1.2/cm and induced with 0.1 mM isopropyl β-D-thiogalactopyranoside (IPTG). The cells were grown for an additional 2 h postinduction, harvested by low-speed centrifugation, washed in buffer A containing 0.5 mg/mL lysozyme (lysis buffer SW; see above), and sonicated.

(*i*) *Purification of Fragment A.* The extract was centrifuged at 10 000 rpm for 30 min, and the supernatant was treated with 0.35% polyethylenimine (PEI; pH 8.0). The PEI precipitate was removed by centrifugation at 10 000 rpm for 30 min. Fragment A and other proteins were precipitated from the PEI supernatant by addition of ammonium sulfate to 65% saturation. The $(NH_4)_2SO_4$ precipitate was dissolved in buffer B and gently heated as described (Agarwal et al., 1991). The precipitate was removed by centrifugation (45 min at 10 000 rpm). The supernatant was dialyzed against buffer C and loaded onto a phosphocellulose column (P-11, 2 cm × 4 cm per 10 g of initial cell pellet). TFIIS fragment A was eluted with a linear NaCl gradient (50–500 mM) in buffer C, exhaustively dialyzed against deoxygenated, deionized, distilled H_2O, and lyophilized. The powder was stored under nitrogen at −70 °C. The peptide was of >98% purity, as determined by SDS–PAGE. Its identity was confirmed by

[2] Amino acids are designated in the text by their three-letter code and in the figures and captions by their one-letter code.

Table I: Chemical Shifts of Assigned ^{15}N and ^{1}H NMR Resonances at pH 6.0

residue	chemicals shifts[a] (25 °C)				
	^{15}N	NH	$C^{\alpha}H$	$C^{\beta}H$	others
T2	95.3	8.12	4.36	4.22	$C^{\gamma}H_3$ 1.18
G3	91.6	8.40	3.97, 3.97		
G4	89.2	8.16	3.80, 4.01		
T5	97.5	8.29	4.35	3.44	$C^{\gamma}H_3$ 1.19
Q6	107.4	8.65	4.80	2.01, 2.12	$C^{\gamma}H$ 2.38, 2.38; $N^{\delta}H$ 6.90, 7.44
T7	98.6	8.93	4.75	4.20	$C^{\gamma}H_3$ 1.22
D8	105.3	8.12	5.04	2.67, 3.02	
L9	100.8	8.09	3.84	0.74, 0.93	$C^{\gamma}H$ 1.24; $C^{\delta}H_3$, 0.55; $C^{\gamma}H_3$ 0.55
F10[b]	97.3	8.19	4.91	3.37, 2.98	$C^{\delta}H$ 7.1; $C^{\epsilon}H$ 6.78; $C^{\zeta}H$ 6.78
T11	97.8	8.42	4.50	3.80	$C^{\gamma}H_3$ 0.88
C12[b]	110.0	8.65	3.92	3.30, 2.96	
G13	97.8	8.35	3.48, 3.74		
K14	103.1	8.85	4.37	1.68, 1.96	$C^{\gamma}H$ 1.33, 1.33; $C^{\delta}H$ 1.56, 1.56; $C^{\epsilon}H$ 3.15, 3.15
C15[b]	100.3	8.62	4.80	2.73, 3.20	
K16	96.8	7.82	4.12	1.97, 2.18	$C^{\gamma}H$ 1.18, 1.32; $C^{\delta}H$ 1.60, 1.60; $C^{\epsilon}H$ 2.90, 2.90
K17	101.4	8.18	4.60	1.90, 2.28	$C^{\gamma}H$ 1.51, 1.57; $C^{\delta}H$ 1.67, 1.67; $C^{\epsilon}H$ 3.05, 3.05
K18	102.4	8.55	5.32	1.64, 1.64	$C^{\gamma}H$ 1.35, 1.35; $C^{\delta}H$ 1.44, 1.44; $C^{\epsilon}H$ 2.98, 2.98
N19[b]	93.6	9.33	5.06	3.07, 2.89	$N^{\gamma}H$ 7.03, 8.53
C20[b]	99.7	8.16	5.45	2.70, 3.05	
T21	97.9	9.16	5.04	3.86	$C^{\gamma}H_3$ 1.04
Y22[b]	104.2	8.92	5.83	2.64, 2.92	$C^{\delta}H$ 6.87; $C^{\epsilon}H$ 6.65
T23	96.6	8.99	4.51	3.94	$C^{\gamma}H_3$ 1.06
Q24	103.8	8.68	5.33	2.02	$C^{\gamma}H$ 2.30, 2.41; $N^{\delta}H$ 6.92, 7.49
V25	103.9	8.93	4.32	2.02	$C^{\gamma}H_3$ 0.90, 0.90
Q26	107.1	8.77	4.58	2.0, 2.14	$C^{\gamma}H$ 2.35, 2.35; $N^{\delta}H$ 6.85, 7.50
T27	98.4	8.46	4.45	4.32	$C^{\gamma}H_3$ 1.23
R28	102.5	8.55	4.42	1.65, 1.76	$C^{\gamma}H$ 2.01, 2.01; $C^{\delta}H$ 3.21, 3.21; $N^{\epsilon}H$ 7.29
S29	96.0	8.08	4.50	3.84, 3.96	
A30	105.7	8.62	4.15	1.41	
D31[b]	95.9	8.16	4.57	2.72, 2.58	
E32	101.1	7.53	4.65	1.93, 2.05	$C^{\gamma}H$ 2.24, 2.24
P33		4.45	2.04, 2.31		$C^{\gamma}H$ 2.23, 2.23; $C^{\delta}H$ 3.64, 3.71
M34	101.3	8.40	4.54	1.92, 1.92	$C^{\gamma}H$ 2.38, 2.54; $C^{\epsilon}H$ 2.32
T35	102.0	8.51	4.45	3.85	$C^{\gamma}H_3$ 0.65
T36	103.9	8.67	4.80	3.18	$C^{\gamma}H_3$ 0.88
F37[b]	106.5	8.94	4.70	2.85, 2.76	$C^{\delta}H$ 7.03; $C^{\epsilon}H$ 7.17; $C^{\zeta}H$ 7.21
V38	104.5	8.84	4.20	−0.40	$C^{\gamma}H_3$ −0.10, 0.04
V39	105.2	8.45	4.67	1.75	$C^{\gamma}H_3$ 0.78, 0.78
C40[b]	108.8	9.34	4.94	3.42, 3.20	
N41[b]	109.9	9.55	4.60	2.88, 2.55	$N^{\gamma}H$ 6.65, 7.16
E42[b]	102.2	9.42	4.47	2.26, 2.18	$C^{\gamma}H$ 2.37, 2.37
C43[b]	101.0	9.35	5.14	2.65, 3.32	
G44	93.1	7.66	3.87, 4.30		
N45[b]	103.0	8.86	4.65	3.23, 3.06	$N^{\gamma}H$ 7.27, 7.37
R46	103.5	7.92	5.65	1.78, 1.88	$C^{\gamma}H$ 1.62, 1.62; $C^{\delta}H$ 3.08, 3.08; $N^{\epsilon}H$ 7.14
W47[b]	103.9	9.05	5.05	3.46, 3.30	$C^{\delta}H2$ 6.93; $C^{\epsilon}H4$ 7.10; $C^{\zeta}H5$ 6.98; $C^{\zeta}H7$, 7.09; $C^{\eta}H6$, 7.09; $N^{\epsilon}H1$ 9.85
K48	99.1	8.50	4.91	1.30, 1.54	$C^{\gamma}H$ 1.02, 1.12; $C^{\delta}H$ 1.42, 1.52; $C^{\epsilon}H$ 2.85, 2.85
F49	101.0	8.63	4.85	2.96	$C^{\delta}H$ 7.04; $C^{\epsilon}H$ 7.04; $C^{\zeta}H$ 7.04
C50	104.3	8.06	4.40	2.85, 2.85	

[a] ^{15}N chemical shifts were measured relative to $^{14}NH_4{}^{15}NO_3$ in D_2O (376.25 ppm); ^{1}H chemical shifts were measured relative to the H_2O resonance, presumed to be 4.78 ppm at pH 6.0 and 25 °C. [b] β protons were stereospecifically assigned and listed in the order of $\beta 2$ and $\beta 3$.

amino acid composition and N-terminal sequencing; the presence of equimolar Zn was verified by atomic absorption spectroscopy.

(*ii*) *Purification of Fragment B.* The extract was centrifuged at 10 000 rpm for 30 min. The supernatant was recentrifuged at 40 000 rpm for 30 min. The second supernatant was loaded (with a flow rate of 5 column vol/h) onto a Ni^{2+}–IDA column (0.2 mL/mg of fragment B; Invitrogen, Inc.), which had been preequilibrated with buffer A. The column was successively washed with 10 column vol each of (i) 20 mM imidazole hydrochloride (pH 7.8) in buffer A, (ii) 40 mM imidazole hydrochloride (pH 7.8) in buffer A, and (iii) 120 mM imidazole hydrochloride (pH 6.0) in buffer D. TFIIS fragment B was eluted with 300 mM imidazole hydrochloride (pH 6.0) in buffer D. Fractions containing fragment B were dialyzed against H_2O, lyophilized, and characterized as above. Purity was >98% as determined by SDS–PAGE.

Circular Dichroism. CD spectra were obtained with an Aviv spectropolarimeter at a peptide concentration of 0.1 mM using a 1-mm path-length cuvette as described (Weiss et al., 1990). Buffers were deoxygenated with N_2 prior to dissolution of the peptide.

NMR Sample Preparation. The peptide–Zn^{2+} powder was dissolved in 0.7 mL of NMR buffer, and the pH (or pD) was adjusted with small aliquots of 0.1 N DCl. Peptide concentrations were 2–3 mM for ^{15}N 3D NMR studies and 8 mM for homonuclear 3D NMR studies. Care was required to avoid acid denaturation (which leads to irreversible precipitation at high protein concentrations) and addition of excess Zn^{2+} (which mediates partial self-association and broadening of NMR resonances).

Homonuclear 2D NMR. The following 2D experiments were performed at 10, 25, and 40 °C with Varian Unity and VXR 500-MHz spectrometers: NOESY (mixing times 100 and 200 ms), DQF-COSY, and TOCSY (mixing time 55 ms). Spectra were obtained in both D_2O and H_2O. Quadrature in both dimensions was achieved by the method of States et al. (1983) with experimental parameters: spectral width 6000 Hz, data size 2K, complex points in t_2 with 300

hypercomplex t_1 increments. Water suppression was achieved by solvent presaturation. Data were processed with a combination of exponential and shifted sine-bell window functions in each dimension. The observed 300 × 2K complex data matrices were zero-filled to 2K × 2K (NOESY and TOCSY) or 1K × 4K (DQF-COSY). $^3J_{\alpha N}$ coupling constants were measured using DQF-COSY in H_2O; distortions in apparent multiplet separation arising from finite line widths were corrected by spectral simulation as described (Redfield & Dobson, 1990; Smith et al., 1991). Chemical shifts are measured in parts per million (ppm) relative to H_2O (presumed to be 4.78 ppm at 25 °C).

Homonuclear 3D NMR. 3D TOCSY-NOESY experiments (Oschkinat et al., 1989a,b; Clore & Gronenborn, 1991) were performed in H_2O and D_2O with the MLEV17 TOCSY mixing scheme. Quadrature detection in ω_1 and ω_2 was obtained by a combination of the methods of time-proportional phase incrementation (TPPI; Marion & Wuthrich, 1983) and States et al. (1983) as described by Bax and co-workers (TPPI–States; Marion et al., 1989). Experimental data consist of 64 × 128 × 512 complex data points; this matrix was extended by linear prediction to 256 × 256 × 512 using software supplied by the vendor. A spectral width of 12 ppm was used in each dimension. Eight scans were collected for each (t_1, t_2) increment with relaxation delay of 1 s. Separate spectra were obtained in D_2O and H_2O; in the latter the solvent resonance was suppressed by presaturation. In each case NOESY and TOCSY mixing times were 200 and 30 ms, respectively. Exponential apodization and shifted sine-bell window functions were applied in each dimension.

Heteronuclear ^{15}N NMR. Experiments were performed with TFIIS fragment B containing uniform ^{15}N enrichment. A 2D HSQC spectrum was obtained with spectral widths of 2000 (ω_1) and 7000 Hz (ω_2) (supplementary material). Water suppression was achieved by using a combination of resonance presaturation and long-pulse methods (Messerle et al., 1989). Experimental data consisted of 512 × 128 complex points and were zero-filled to 2048 × 1024 complex points. 3D NOESY-HMQC spectra (Clore & Gronenborn, 1991) were collected with spectral widths of 7000 Hz in both ω_1 and ω_3 and with 1500 Hz in ω_2. A total of 32 × 128 × 512 data points were obtained. The data were extended by linear prediction to 128 × 256 × 512. In each case the NOESY mixing time was 200 ms. Quadrature detection in ω_1 and ω_2 was obtained using the TPPI–States method.

Structure Calculation. NOE and *J*-coupling (dihedral angle) restraints were used for molecular modeling as described (Kochoyan et al., 1991). Distance geometry/simulated annealing (DG/SA) calculations were performed with the program DGII (Havel, 1991). Possible hydrogen bonds in DG structures were inferred from preliminary DG/SA models using the analysis facility of X-PLOR (A. T. Brunger, Yale University). Only those peptide and amide-sulfur hydrogen bonds consistent with slowly exchanging amide resonances in D_2O were incorporated as additional distance restraints in refined DG/SA calculations as described (Qian & Weiss, 1992). Other hydrogen bonds are considered as stereochemically permitted but speculative. To avoid systematic errors in DG/SA calculations due to use of hard-sphere atoms, lower bounds between amide protons and possible hydrogen-bond acceptors were reduced as described (Akke et al., 1992; Qian & Weiss, 1992).

Database Search. Analogous Cys_4 sequences were obtained with a template and homology matching protocol. The template Cys-X_2-Cys-X_{6-9}-h-X_{14}-Cys-X_2-Cys-X_3-h (in which h designates a hydrophobic residue) was created based on inspection of TFIIS-related sequences. Matching sequences were obtained from the SWISS-PROT (version 24.0), PIR (version 35.0), and GenBank (release 74.0 with cumulative daily updates) databases. The initial set was reduced by elimination of (i) classical (TFIIIA-like) Zn finger sequences, (ii) Cys_4 sequences containing multiple prolines within the putative β-strands, and (iii) sequences of proteins whose functions are unrelated to nucleic acid transactions. The remaining sequences were then used as the basis for a generalized homology search (BLASTP program; Altschul et al., 1990) of the same three databases. The second search enabled identification of sequences that deviate from the original template (e.g., sequences containing Cys-X_4-Cys elements). Multiple alignment of the final set of sequences was obtained using the CLUSTAL program (Higgins & Sharp, 1988, 1989).

Homology Modeling. A structure of an analogous Cys_4 domain in human TFIIE was built using the homology-modeling method of Havel (1990). The Cys_4 sequences of TFIIS and TFIIE were aligned as described (Higgins & Sharp, 1988, 1989). Optimal alignment predicted that the TFIIE domain contains a three-residue deletion between the second cysteine and first β-strand (TFIIS peptide residues 11–13; see Figure 15). DG/SA models of the TFIIE domain were obtained based on 296 distance and 52 dihedral restraints as described under Results. Analogous Zn coordination, β-sheet, and amide-sulfur hydrogen bonds were assumed. The structures were refined by restrained molecular dynamics and energy minimization using the same protocol as employed for NMR structures (Clore et al., 1985; Kochoyan et al., 1991).

RESULTS

Our results are presented in four sections. Section I contains an overview of TFIIS Zn coordination and metal-dependent protein folding. 3D NMR studies are presented in section II. Distance geometry/simulated annealing calculations and molecular modeling of the TFIIS nucleic acid binding domain are presented in section III. A database search for analogous Zn-binding motifs is described in section IV. The three-dimensional structure of one analogue, a putative Zn-binding site in basal transcriptional initiation factor TFIIE, is predicted based on homology modeling (Havel, 1990).

(I) Zn Coordination and Overall Folding Properties

Zn Coordination. The primary structure of the C-terminal domain of human TFIIS contains six cysteines (Cys12, Cys15, Cys20, Cys40, Cys43, and Cys50; using the peptide numbering scheme given in Figure 1); there are no histidines in the native sequence (Yoo et al., 1991). The stoichiometry of Zn^{2+} binding, as determined by atomic absorption spectroscopy, is one Zn per polypeptide. XAFS experiments demonstrate that Zn^{2+} coordination is accomplished only by cysteines (Agarwal et al., 1991). Which cysteines coordinate Zn^{2+} is suggested by inspection of TFIIS homologues: Cys12, Cys15, Cys40, and Cys43 are invariant among species whereas C20 and C50 are not. A homologue in *S. cerevisiae*, for example, contains the substitutions Cys20Val and Cys50Ser (Clark et al., 1991; Nakanishi et al., 1992). Introduction of these substitutions into human TFIIS does not affect its proteolytic stability in *E. coli* nor its ability to stimulate RNAP-II transcriptional elongation (C. J. Jeon, H.-S. Yoon, and K. Agarwal, manuscript in preparation). In contrast, substitution of an invariant cysteine (Cys12Ser, Cys15Ser, Cys40Ser, or Cys43Ser) leads to undetectable levels of bacterial expression, presumably due to efficient proteolysis of an unfolded or misfolded polypeptide (Agarwal et al., 1991). We adopt as

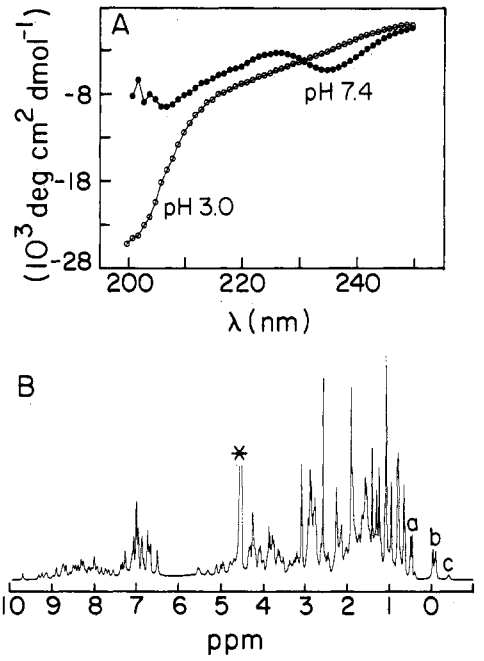

FIGURE 2: (A) CD spectra of folded (●; pH 7.4) and acid-denatured (○; pH 3.0) TFIIS-derived peptide (fragment A) at 0 °C in the presence of equimolar $ZnCl_2$. The peptide concentration was 100 μM. (B) One-dimensional ^1H NMR spectrum of peptide–Zn^{2+} complex (fragment A) at 25 °C in 50 mM deuterated Tris-HCl (pH 6.0) in 90% H_2O/10% H_2O. The peptide concentration was 5 mM. The asterisk indicates residual solvent resonance following presaturation.

a working hypothesis a model in which invariant cysteines Cys12, Cys15, Cys40, and Cys43 coordinate Zn^{2+} (as depicted in Figure 1A). This hypothesis is directly tested in section III by distance-geometry calculations.

Circular Dichroism. The far-UV CD spectrum of the TFIIS fragment A–Zn^{2+} complex at pH 7.4 is shown in Figure 2A (filled circles). The spectrum indicates the presence of nonrandom structure but is not readily decomposed into subspectra assigned to elements of secondary structure. There is no similarity between the TFIIS spectrum and those of classical Zn fingers, which exhibit partial α-helical character (Parraga et al., 1988; Weiss et al., 1990). Absence of the α-helix in the TFIIS domain would be in accord with neural-net prediction of secondary structure (Hulley & Karplus, 1989): such calculations predict a predominance of β-sheet. Extended structure is predicted between residues 8–11, 21–28, and 35–40 (neural-net score 0.2–0.8).

Corresponding CD spectra of the apopeptide could not be obtained at pH 5-8 due to macroscopic precipitation, thus precluding investigation of Zn-dependent structural changes. However, a pH-dependent unfolding transition is observed under acidic conditions that is analogous to those of classical Zn fingers (Weiss et al., 1990). The CD spectrum of TFIIS fragment A at pH 3.0 (open circles in Figure 2A) is as expected for a random coil and is not affected by the addition of Zn^{2+}. Presumably, destabilization of Zn^{2+} coordination occurs under acidic conditions due to protonation of Cys thiolate. These data suggest but do not establish that the folded structure of the fragment is Zn-dependent at neutral pH.

1D NMR Studies. The ^1H NMR spectrum of the TFIIS fragment A–Zn^{2+} complex in H_2O is shown in Figure 2B. Marked dispersion of chemical shifts provides evidence of a stably folded structure. Resonance line widths are consistent with the presence of a monomer as previously inferred from gel filtration chromatography (Agarwal et al., 1991). In the absence of a reducing agent, however, progressive line broadening is observed over 48 h, presumably due to the formation of intermolecular disulfide bonds between cysteines not coordinated to Zn^{2+} (Cys20 and/or Cys50; Figure 1A). Such oligomerization is readily reversed by the addition of 10 mM deuterated DTT and storage of samples under anaerobic conditions (see Materials and Methods).

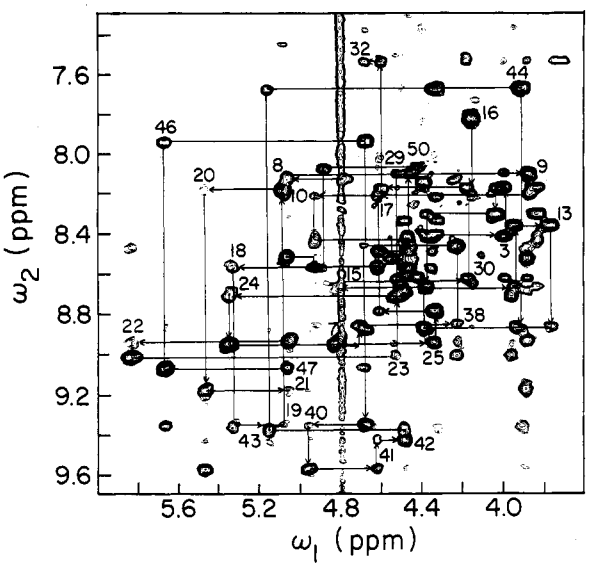

FIGURE 3: Fingerprint region of homonuclear two-dimensional NOESY spectrum of peptide–Zn^{2+} complex (fragment B); sequential assignment is as indicated. Conditions were as described in the legend to Figure 2B. The mixing time was 200 ms.

Identical oxidation-dependent oligomerization is observed with TFIIS fragment B. Its 1D ^1H NMR spectrum is similar to that of fragment A in resonance dispersion and detailed pattern of upfield-shifted aliphatic resonances (supplementary material). The line widths of such resonances are increased in the larger fragment; as expected, random-coil resonances assigned to the additional N-terminal residues are observed, including degenerate H_ϵ and H_δ aromatic resonances from the N-terminal His_6 "tag". The detailed NMR analysis presented below will focus on fragment B because it can be isotopically labeled and purified in greater yield than can fragment A. The additional nonnative N-terminal residues are disordered and will not be considered further.

(II) 3D NMR Studies and Resonance Assignment

Sequential Resonance Assignment. The "fingerprint" region of the conventional 2D NOESY spectrum (Wuthrich, 1986) of fragment B is shown in Figure 3. Partial sequential $d_{\alpha N}$ connectivities are outlined. Resolution is limited in this region, however, rendering ambiguous assignment of key NOEs. Such limitations are readily circumvented by the application of homonuclear and heteronuclear ^{15}N 3D NMR methods. In Figure 4A is shown, for example, a crowded subsection of the NOESY fingerprint region. In panels B–F are shown successive NOESY ω_2–ω_3 planes of a 3D TOCSY-NOESY spectrum obtained in H_2O. Assignment of sequential NOEs is simplified by filtering via the ω_1 TOCSY frequency (Oschkinat et al., 1989a,b). In this example, NOEs are observed between amide resonances (ω_3) and those H_α resonances (ω_2) which belong to a spin system labeled by TOCSY frequency ω_1.

A complementary method to resolve sequential NOEs is provided by ^{15}N-editing. For this purpose, fragment B was uniformly labeled by expression in ^{15}N-enriched minimal medium (labeling efficiency >98%; see Materials and Methods). An HSQC spectrum of the labeled protein is provided

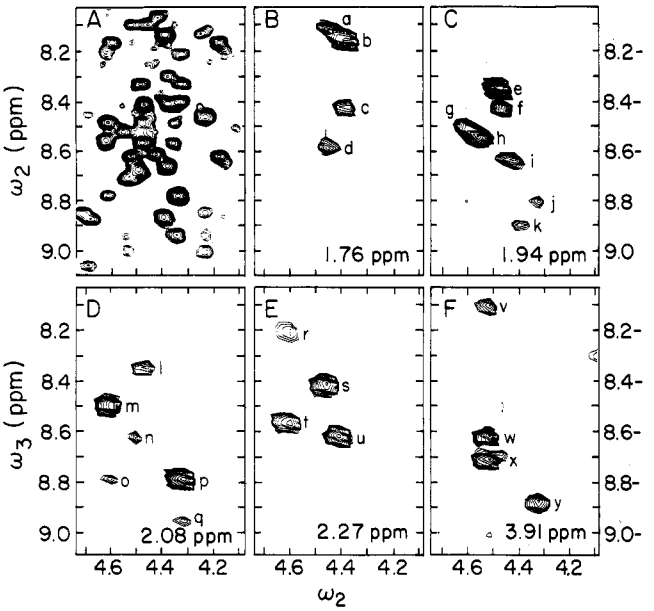

FIGURE 4: (A) Portion of 2D-NOESY fingerprint region with severe resonance overlapping (see Figure 3). (B–F) Corresponding NOESY cross-planes of homonuclear 3D TOCSY-NOESY spectrum at ω_1 = 1.76 ppm (B), ω_1 = 1.94 (C), ω_1 = 2.08 (D), ω_1 = 2.27 (E), and ω_1 = 3.91 (F). Assignments of 3D cross-peaks in panel B: (a) R28H_α-S29H_N, (b) K1H_α-T2H_N, (c) K1H_N-K1H_α, and (d) R28H_N-R28H_α. Assignments of 3D cross-peaks in panel C: (e) M2H_α-A1H_N, (f) P33H_α-M34H_N, (g) Q26H_α-T27H_N, (h) M34H_α-T35H_N, (i) K14H_α-C15H_N, (j) V25H_α-Q26H_N, and (k) K14H_N-K14H_α. Assignments of 3D cross-peaks in panel D: (l) $M_{-2}H_\alpha$-$A_{-1}H_N$, (m) Q26H_α-$T_{27}H_N$, (n) S29H_α-A30H_N, (o) Q26H_N-Q26H_α, (p) V25H_α-Q26H_N, and (q) V25H_N-V25H_α. Assignments of 3D cross-peaks in panel E: (r) K17H_N-K17H_α, (s) P33H_α-M34H_N, (t) K17H_α-K18H_N, and (u) $M_{-2}H_N$-$M_{-2}H_\alpha$. Assignments of 3D cross-peaks in panel F: (v) S29H_N-S29H_α, (w) S29H_α-A30H_N, (x) T23H_α-Q24H_N, and (y) G44H_α-N45H_N. For spectrum A, horizontal and vertical axes are ω_1 and ω_2, respectively; for spectra B–F horizontal and vertical axes are ω_2 and ω_3, respectively.

as supplementary material. As observed in other systems, there is no apparent correlation between H_N and ^{15}N chemical shifts, facilitating resolution of individual NOEs in a 3D NOESY-HMQC experiment. Representative ω_1–ω_3 strips arranged in order of the amino acid sequence are shown in Figure 5A; sequential NOEs are outlined. Such relationships are extracted from individual ^{15}N-edited ω_1–ω_3 NOESY plane; one such plane is shown in Figure 5B. Complete resonance assignment was obtained by integrated analysis of homonuclear and heteronuclear 3D NMR data. ^1H and ^{15}N chemical shifts are given in Table I. Figure 6 summarizes sequential and medium-range NOEs, $^3J_{HN}$ coupling constants, selected χ_1 angles based on stereospecific assignment of β protons (asterisks in Table I), and location of slowly exchanging amide resonances at pD 6.0 and 25 °C (data not shown).

Analysis of Secondary Structure. Figure 7 shows a diagonal plot of interresidue NOEs; main-chain contacts are summarized at the lower right. Formation of a three-stranded antiparallel β-sheet (residues 20–25, 35–40, and 45–48) is implied by two strings of cross-peaks perpendicular to the diagonal. Observed interstrand contacts in the β-sheet are indicated by arrows in Figure 8. Three β-turns are also seen (14–16, 30–32, and 42–44), which include the Zn-bound cysteates (see below). Medium- and long-range NOEs occur between segments 5–11 and 15–20. Although the N-terminal region is essentially extended, it cannot be assigned to a regular secondary structure. Unlike the classical Zn finger (Parraga et al., 1988; Lee et al., 1989; Omichinski et al., 1990; Pavletich & Pabo, 1991), glucocorticoid-related Cys_4 domains (Schwabe et al., 1990; Luisi et al., 1991), and the Gal4 Cys_6 domain

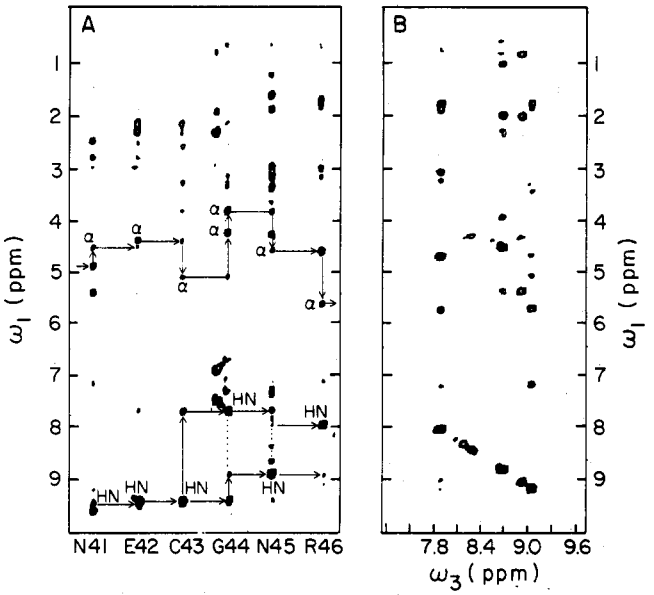

FIGURE 5: Heteronuclear 3D ^{15}D–^1H NOESY-HMQC spectrum of uniformly ^{15}N-labeled peptide–Zn^{2+} complex at 25 °C. (A) Alignment of selected ^{15}N planes in sequence order to illustrate the sequential assignment of ^{15}N-edited fingerprint region. (B) Representative NOESY plane of 3D spectrum. The NOESY mixing time was 200 ms, and the protein concentration was 2 mM.

(Marmorstein et al., 1992; Baleja et al., 1992; Kraulis et al., 1992), no α-helix is observed.

Tertiary Structure. In addition to main-chain NOEs, medium- and long-range interactions are observed that define the structure of the Zn-binding site and side-chain packing interactions.

(*i*) *Zn-Binding Site.* NOEs are observed from Cys12 to Cys15 and Cys40, which reflect tetrahedral Zn^{2+} coordination (Cys12-H$_{\beta2}$/Cys15-H$_N$ and Cys12-H$_\alpha$/Cys40-H$_{\beta1,2}$). Short- and medium-range contacts indicate that the two Cys-X_2-Cys regions form β-turns. Additional contacts include Cys12-H$_{\beta2}$/Lys16-H$_N$, Cys40-H$_N$/Asn45-H$_N$, and Cys40-H$_N$/Arg46-H$_{\beta1,2}$. NOEs and dihedral restraints in the Zn-binding site also predict two sets of bifurcating amide-sulfur hydrogen bonds as described below (section III).

(*ii*) *Side-Chain Contacts.* Particularly prominent are long-range NOEs involving the aromatic rings of Phe10, Tyr22, Phe37, and Trp47. Contacts are observed from Phe10 in the N-terminal region to β-strand 1 (Tyr22) and β-strand 2 (Thr36 and Val38). The aromatic ring of Tyr22 contacts side chains of three additional N-terminal residues (Thr5, Thr7, and Leu9) and two residues in β-strand 2 (Thr36 and Val38). The aromatic ring of Phe37 in β-strand 2 contacts side chains in β-strands 1 (Val25) and 3 (Lys48). The aromatic ring of Trp47 contacts side chains in the Zn-binding site (Cys12 and Cys40), in β-strand 2 (Val38 and Val39), and within β-strand 3 (Asn45) as described below.

Resolution of aromatic-methyl NOEs is facilitated by the analysis of the homonuclear TOCSY-NOESY 3D spectrum recorded in D_2O. An example is provided in Figure 9. Although the two methyl resonances of Thr5 and Thr7 have nearly identical chemical shifts (Table I), their distinct NOEs to ortho resonance of Tyr22 are resolved in different ω_2–ω_3 planes of the 3D spectrum. Superimposed cross-peaks in the 2D NOESY spectrum (cross-peak a in Figure 9A) are edited by the resolved H$_\beta$ frequencies of Thr5 and Thr7 in ω_1 (3.44 and 4.19 ppm, respectively). These are shown as cross-peaks a′ and a″ in panels B and C. A table of NOE contacts is provided as supplementary material.

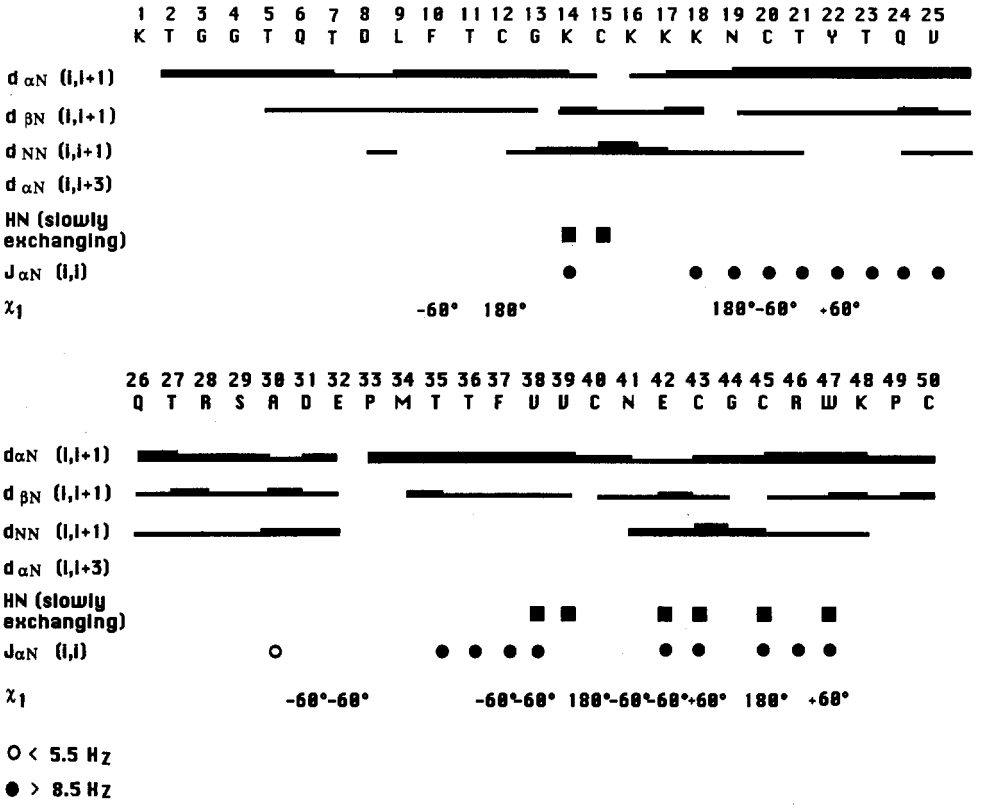

FIGURE 6: Summary of sequential and medium-range NOEs; symbols ($d_{\alpha N}$, d_{NN}, etc.) are as defined by Wuthrich (1986). Relative intensities of NOEs are schematically indicated by the thickness of the line; χ_1 values are indicated for the residues for which stereospecific assignment of β resonances was obtained (asterisks in Table I). Slowly exchanging amide resonances at 25 °C are indicated.

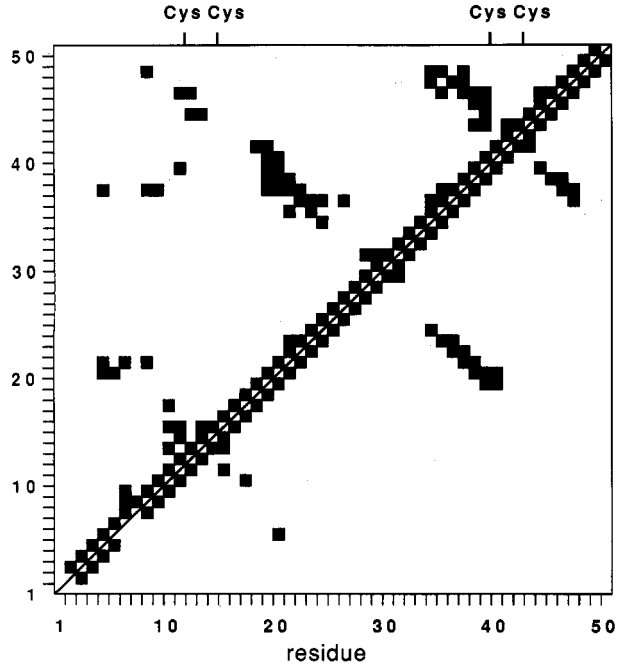

FIGURE 7: Diagonal plot of interresidue NOE cross-peaks. Points above diagonal represent NOEs between any two protons; points below diagonal designate contacts between main-chain protons only. Positions of Zn^{2+}-bound cysteines are indicated.

(III) Solution Structure

The program DG-II (Havel, 1991) was used to generate the DG/SA models of the TFIIS nucleic acid binding domain (peptide residues 5–50). Twenty structures were obtained based on 934 restraints as summarized in Table II. Tetrahedral coordination of Zn^{2+} by Cys12, Cys15, Cys40, and Cys43 was explicitly imposed: the four S_γ–Zn bond distances

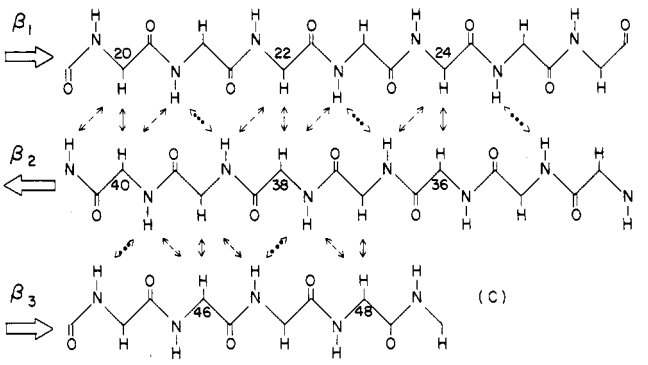

FIGURE 8: Three-stranded antiparallel β-sheet is defined in the peptide–Zn^{2+} complex by interstrand backbone NOEs: solid arrows designate $d_{\alpha\alpha}(i,j)$ contacts; dashed and dotted arrows designate $d_{\alpha N}(i,j)$ and $d_{NN}(i,j)$ contacts, respectively.

were constrained to be 2.2–2.4 Å, and six S_γ–S_γ distances were constrained to be 3.6–4.0 Å. In refined structures, hydrogen-bond restraints were introduced based on slowly exchanging amide resonances (see below). Of the 20 structures, 14 had a residual error function of less than 1.1 kcal/mol and were subjected to restrained molecular dynamics without explicit charges (Clore et al., 1985; Kochoyan et al., 1991). In this set there were no NOE violations >0.4 Å.

Main-Chain Conformation. An ensemble of backbone structures (residues 5–50) is shown in Figure 10A; a corresponding ribbon model is shown in Figure 1B. The structure is most precisely defined between residues 5–26 and 35–48; in these regions the main-chain RMS deviation (relative to mean coordinate positions) is 0.40 Å and side-chain RMS deviation is 0.89 Å (Table II). The structure is elongated: the central β-sheet contains an internal Zn-binding site at one end and an exposed disordered loop at the other. The Zn-binding site consists of well-defined β-turns whose configuration is stabilized by contacts with side chains in the β-sheet;

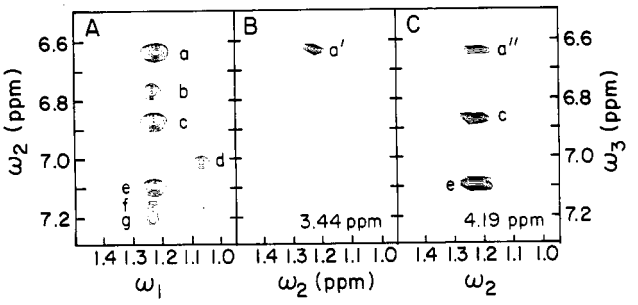

FIGURE 9: Homonuclear 3D TOCSY-NOESY spectrum in D_2O. (A) Baseline 2D NOESY spectrum (aliphatic–aromatic region) in D_2O. Resonance overlap occurs in the aliphatic dimension (ω_2). (B and C) NOESY ω_2–ω_3 cross-planes at ω_1 = 3.44 ppm (B) and ω_1 = 4.19 ppm (C). Assignment of cross-peaks: (a′) T5(γ-CH_3)-Y22(ortho), (a″) T7(γ-CH_3)-Y22(ortho), (b) T7(γ-CH_3)-F10(para), (c) T7(γ-CH_3)-Y22(para), (d) F37(ortho)-T23(γ-CH_3), (e) T7(γ-CH_3)-F10(ortho), (f) F37(para)-T27(γ-CH_3), and (g) F37(para)-T27(γ-CH_3).

the Sγ atoms of Cys20 and Cys50 are too distant from Zn^{2+} to act as ligands (6 and 20 Å, respectively). The loop occurs between the first and second β-strands and contains the fingertip (Figure 1). One representative DG/RMD model is shown in stereo in Figure 10B with selected side chains. The precision of the DG/RMD ensemble is shown by residue in Figure 11A as RMS deviations among main-chain and side-chain atoms (open and closed circles, respectively). Although the side chain of Phe49 (β-strand 3) exhibits prominent long-range NOEs, its position is not well-defined due to degeneracy of ring resonances (Table I). The distribution of main-chain ϕ and ψ dihedral angles in the ensembles (Ramachandran plot) is as expected in the ordered regions as shown in Figure 12. The 3D profile score, defined by Bowie et al. (1991) as a measure of the fit between sequence and structure, is 16.1 ± 1.4 (residues 1–50); this value is in accord with a compatible model.

Table II: NMR Restraints and Statistical Parameters

A. Summary of Restraints			
NOE upper bonds	637	NOE lower bonds[a]	236
intrareside	162	dihedral angle ϕ	20
sequential	216	dihedral angle χ_1	15
long-range	211	Zn coordination[b]	10
medium-range	48	hydrogen bonds[c]	16
B. RMS Deviations (Excluding Disordered Regions[d])			
main chain (Å)	0.40	side chains (Å)	0.89
C. Average Restraint Violations[e]			
upper-bound NOE (Å)	0.028	dihedral angles (deg)	1.2
D. Deviations from Ideal Covalent Geometry			
bond lengths (Å)	0.01	bond angles (deg)	2.92
E. Empirical Energy Function (kcal/mol)			
NOE residual violation[f]	21.2	dihedral angles	160.4
dihedral restraint violation[f]	0.1	improper dihedral angles	3.9
bonds	13.1	van der Waals	−224.1
hydrogen bonds	−47.8	electrostatic	3.1
angles	171.1	total	98.6

[a] A lower bound of 3.0 Å was imposed for weak NOEs (restraint range 3.0–4.3 Å). [b] Tetrahedral Zn^{2+} coordination was imposed as follows: Zn–Sγ and Sγ–Sγ distances were restrained to be 2.2–2.4 and 3.6–4.0 Å, respectively. [c] Two distance restraints were imposed per hydrogen bond; in the case of peptide hydrogen bonds these were H···O=C (1.8–2.4 Å) and NH···O=C (2.7–3.3 Å); in the case of amide-sulfur hydrogen bonds these were HN···Sγ (2.2–2.8 Å) and HN···Sγ (3.1–3.7 Å). [d] The ensemble was aligned according to the C$^\alpha$ positions of residues 5–25 and 35–48; RMS values shown correspond to these regions as calculated against average coordinates. [e] NOE restraint violations represent RMS upper-bound violations; lower-bound violations were 0.017 Å. [f] The NOE- and dihedral-related force constants were set to be 40 kcal/Å^2 and 40 kcal/rad^2, respectively.

Control Calculations. Tetrahedral Zn^{2+} coordination by Cys12, Cys15, Cys40, and Cys43 was explicitly imposed in the above model. Although this assumption is supported by

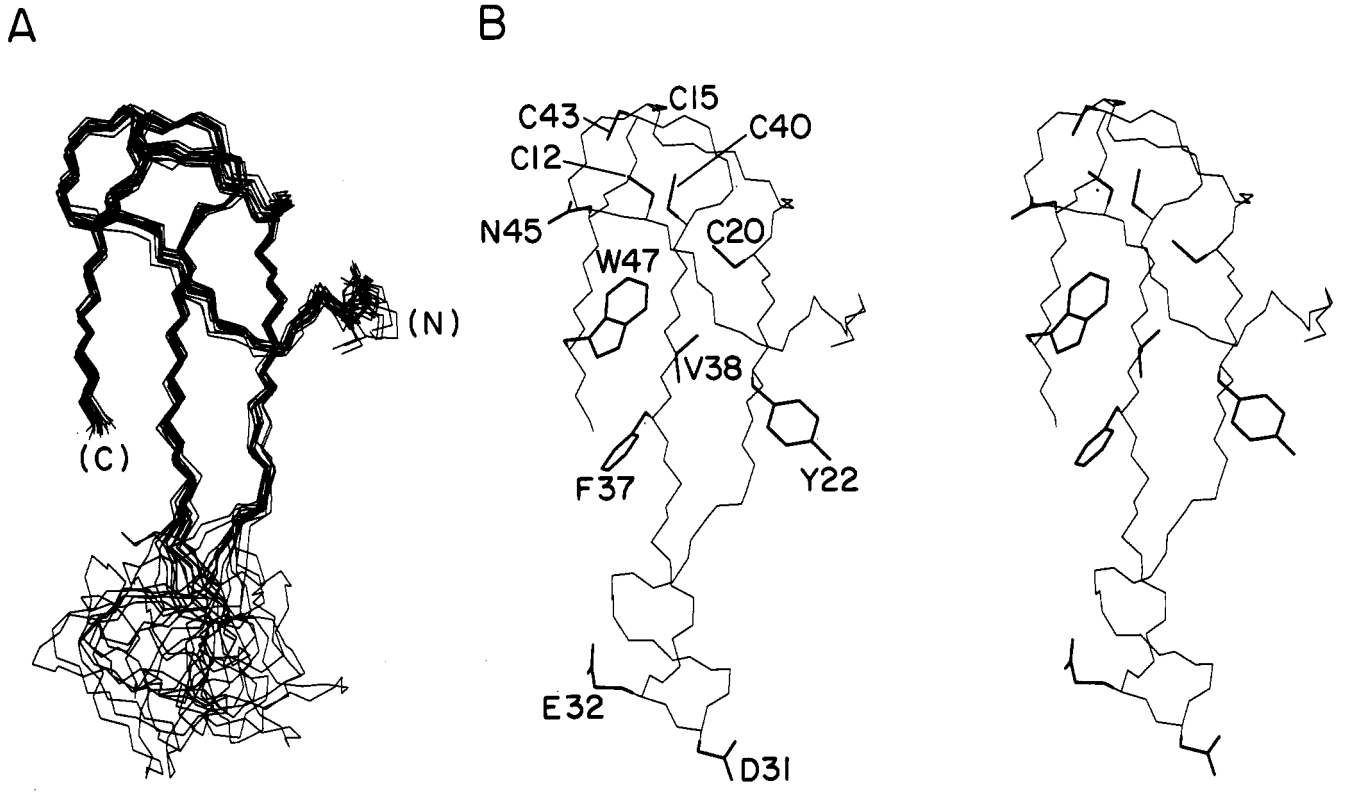

FIGURE 10: (A) Ensemble of 12 DG/SA models, aligned with respect to the main-chain atoms of well-defined residues (see Table II). (B) Stereo pair of one DG/RMD model showing positions of well-ordered side chains (C12, C15, C20, Y22, F37, V38, C40, C43, N45, and W47); disordered side chains D31 and E32 are also shown. Local environments of well-ordered side chains are shown in Figures 13 and 14. Figure is adapted from Qian et al. (1993).

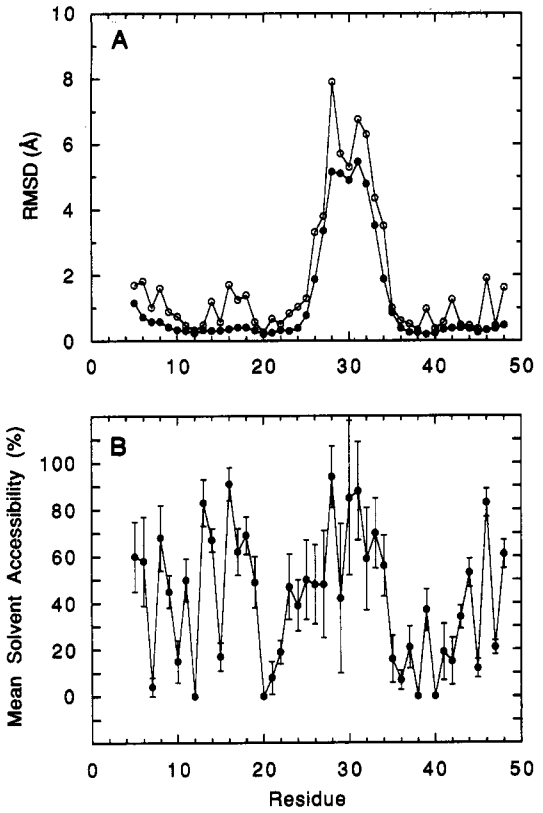

FIGURE 11: (A) RMSD about mean coordinate position shown by residue for main-chain (●) and side-chain atoms (○). (B) Relative solvent accessibilities of side chains as calculated using the Lee and Richards algorithm implemented in XPLOR (probe radius 1.6 Å). Percent values are defined relative to the accessibility of a given side chain (X) imbedded in Gly-Gly-X-Gly-Gly pentapeptide, assumed to be in an extended conformation. 100% indicates fully exposed, and 0% indicates fully buried. Values (●) represent an ensemble average; error bars represent ±1 RMSD.

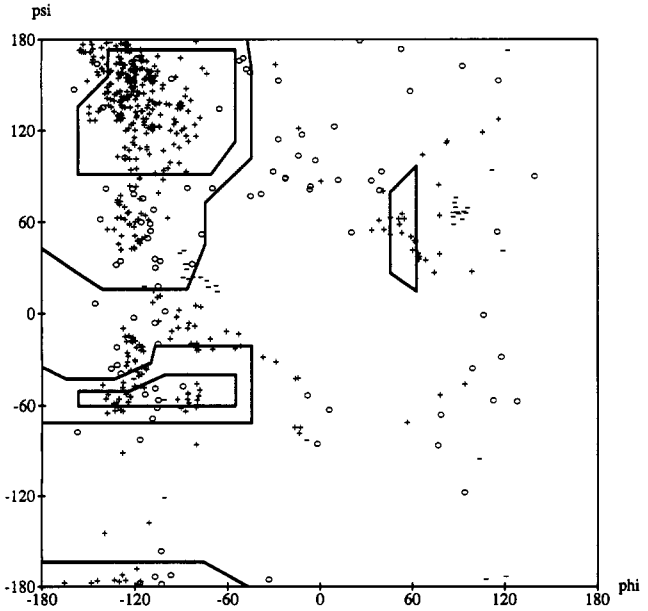

FIGURE 12: Ramachandran plot depicting ensemble distribution of ϕ–ψ backbone angles (residues 4–48). ϕ–ψ angles of non-glycine residues in the ordered region (residues 4–25 and 35–48) are indicated by "+"; those of glycine residues are indicated by "–". ϕ–ψ angles in the disordered loop (residues 26–34) are not well-defined (○).

inspection of homologous sequences and by site-directed mutagenesis (section I), there are in fact 15 combinations of four ligands that may be selected from a set of six cysteines. To test the consistency between the assumed coordination and observed NOEs, two control calculations were performed.

Table III: Geometry of Possible Zn-Binding Sites in Control DG/RMD Models[a]

residue		distance (Å)	residue		distance (Å)
		Control A			
C12	Sγ-Zn	1.9 ± 0.6	C40	Sγ-Zn	2.5 ± 0.2
C15	Sγ-Zn	2.8 ± 0.3	C43	Sγ-Zn	2.6 ± 0.9
C20	Sγ-Zn	6.0 ± 0.4	C50	Sγ-Zn	20.0 ± 2.1
		Control B			
C12	Sγ-Zn	2.0 ± 0.2	C40	Sγ-Zn	2.2 ± 0.2
C15	Sγ-Zn	3.4 ± 0.4	C43	Sγ-Zn	4.5 ± 1.2
C20	Sγ-Zn	4.1 ± 0.2	C50	Sγ-Zn	18.8 ± 2.0
		Control C			
C12	Sγ-Zn	1.6 ± 0.6	C40	Sγ-Zn	2.2 ± 0.4
C15	Sγ-Zn	3.1 ± 0.7	C43	Sγ-Zn	3.7 ± 0.9
C20	Sγ-Zn	4.8 ± 0.5	C50	Sγ-Zn	19.1 ± 2.0
		Control D			
C12	Sγ-Zn	2.2 ± 0.5	C40	Sγ-Zn	1.5 ± 0.1
C15	Sγ-Zn	4.1 ± 0.5	C43	Sγ-Zn	3.7 ± 0.9
C20	Sγ-Zn	4.4 ± 0.4	C50	Sγ-Zn	18.6 ± 2.0
		Control E			
C12	Sγ-Zn	2.2 ± 0.6	C40	Sγ-Zn	1.5 ± 0.2
C15	Sγ-Zn	3.4 ± 0.4	C43	Sγ-Zn	3.6 ± 0.8
C20	Sγ-Zn	4.7 ± 0.4	C50	Sγ-Zn	19.3 ± 2.0

[a] Control DG/RMD models were calculated in the absence of explicit Zn-related restraints. Values shown represent the average (±RMSD) of 10 DG/RMD structures. Five separate control calculations were performed. In calculation A, the Zn position was assumed to be at the center of mass of the sulfur atoms of residue 12, 15, 40, and 43 (the native Zn-binding site; indicated by boldface). In calculations B–E, alternative binding sites involving Cys20 were systematically evaluated. The Zn was placed at the center of mass of the following sulfurs (boldface): (B) residues 12, 15, 20, and 40; (C) residues 12, 15, 20, and 43; (D) residues 12, 20, 40, and 43; and (E) residues 15, 20, 40, and 43.

Table IV: Sγ-Zn-Sγ′ Bond Angles

angle	restrainted ensemble[a]	control ensemble[b]
Sγ(12)-Zn-Sγ(15)	107.6 ± 1.7	88.9 ± 14.4
Sγ(12)-Zn-Sγ(40)	96.8 ± 3.2	104.2 ± 6.7
Sγ(12)-Zn-Sγ(43)	105.6 ± 4.5	114.1 ± 18.3
Sγ(15)-Zn-Sγ(40)	116.4 ± 5.3	126.9 ± 7.7
Sγ(15)-Zn-Sγ(43)	113.2 ± 4.0	121.8 ± 1.4
Sγ(40)-Zn-Sγ(43)	114.1 ± 4.3	98.6 ± 6.4

[a] Tetrahedral Zn^{2+} coordination was imposed as follows: Zn-Sγ and Sγ-Sγ distances were restrained to be 2.2–2.4 and 3.6–4.0 Å, respectively; Sγ-Zn-Sγ′ bond angles were not restrained (see Materials and Methods). [b] In the control ensemble, no Zn-related restraints were employed. The center of mass of Sγ(12), Sγ(15), Sγ(40), and Sγ(43) was used as the presumed Zn position (see footnote to Table III).

(i) DG/SA structures were calculated without imposition of explicit Zn coordination. Such structures are similar to the original model. In particular, the Sγ atoms of Cys12, Cys15, Cys40, and Cys43 are near the presumed position of Zn^{2+} (control A in Table III) with geometry compatible with tetrahedral coordination (Table IV). In contrast, the Sγ atoms of Cys20 and Cys50 cannot readily be fit into a Zn-binding site (e.g., controls B–E in Table III). As expected, the control ensemble is less precise than the original model in the neighborhood of the Zn^{2+}-binding site (data not shown). (ii) Imposition of alternative Zn^{2+} coordination (e.g., Cys12, Cys15, Cys40, and Cys50) leads to large NOE restraint violations. Taken together, these control calculations demonstrate that the NMR data are sufficient in themselves to identify which cysteines coordinate Zn^{2+}. An analogous DG-based procedure to establish disulfide pairing has recently been described (Adler et al., 1993).

Assignment of Hydrogen Bonds. Putative hydrogen bonds were identified in DG/RMD structures using the analysis facility of X-PLOR (Tables V and VI). Some (but not all)

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

donor		acceptor		no. of structures involved	energy (kcal/mol)	distance (Å)	angle [Δ(deg)][b]	class
5	NH	OG_1	21	8	−2.6	3.1	18	side chain
6	NH	OG_1	21	8	−1.2	2.9	41	side chain
7	NH	O=C	20	9	−2.2	3.1	16	tertiary
12	NH	O=C	17	14	−2.2	3.2	20	tertiary
16	NH	O=C	12	13	−0.3	3.3	49	tertiary
19	NH	OE_2	42	7	−0.5	3.3	32	side chain
21	NH	O=C	39	14	−2.7	2.9	17	β-sheet (β1–β2)
23	NH	O=C	37	14	−2.7	3.0	10	β-sheet (β1–β2)
24[c]	$N_\epsilon H_2$	OG_1	36	10	−2.2	3.3	22	side chain
25	NH	O=C	35	11	−1.4	3.4	30	β-sheet (β1–β2)
35	NH	O=C	25	11	−1.3	3.1	30	β-sheet (β1–β2)
37	NH	O=C	23	14	−2.3	3.1	18	β-sheet (β1–β2)
38	NH*	O=C	47	14	−2.9	2.9	14	β-sheet (β2–β3)
39	NH*	O=C	21	14	−3.2	3.0	9	β-sheet (β1–β2)
40	NH*	O=C	45	14	−2.3	2.8	14	β-sheet (β2–β3)
41	$N_\delta H_2$	OG_1	21	10	−1.5	3.0	36	side chain
41	NH	O=C	19	12	−0.7	3.5	35	β-sheet (β2–β3)
45	NH	O=C	40	9	−0.0	3.8	43	β-sheet (β2–β3)
47	NH*	O=C	38	14	−3.0	3.0	10	β-sheet (β2–β3)

[a] Presumptive hydrogen bonds are listed only if present in at least 50% of the 14 DG/RMD models. Carbonyl oxygen acceptors are designated O=C. Asterisks designate slowly exchanging amide proton resonances in D_2O. [b] Δ(deg) is defined as deviation from linearity. [c] The side chain C=O of Gln24 is also positioned to receive a hydrogen bond from the phenolic OH of Tyr22; this hydrogen bond was not predicted by XPLOR because the orientation of the para OH is random in the DG/RMD ensemble.

FIGURE 13: In each Cys-X_2-Cys element, the N-terminal cysteate participates in bifurcating sulfur-amide hydrogen bonds as shown for Cys40 (dotted lines). The C-terminal cysteate (Cys43) does not participate in hydrogen bonding. One representative model is labeled at the left; the corresponding DG/RMD ensemble is shown in stereo at the right. The two sulfur atoms are shown as filled circles; oxygen, carbon, nitrogen, and hydrogen positions are indicated by open circles (radii O>C=N>H). The side chains of Asn41, Glu42, and Asn45 are also shown.

such hydrogen bonds correspond to slowly exchanging amide resonances in D_2O: (i) Val38, Val39, Cys40, and Trp47 (asterisks in Table V), and (ii) Lys14, Cys15, Glu42, and Cys43 (asterisks in Table VI, set A). The former occur in the β-sheet (see Figure 8). The latter correspond to bifurcating sulfur-amide hydrogen bonds in the Zn-binding site ($S\gamma(i) - H_N(i+2)$ and $S\gamma(i) - H_N(i+3)$); only the N-terminal Cys S_γ in each Cys-X_2-Cys element is involved (Cys12 and Cys40). Figure 13 shows the structure of the Cys40-X_2-Cys43 element and Cys40 sulfur-amide hydrogen bonds (dotted lines). The S_γ atoms of Cys15 and Cys43 are, in contrast, more distant from possible hydrogen-bond donors. The nearest amide protons (of Lys17 and Asn45, respectively) are ≈4 Å away (Table VI, set B and Figure 13). In refined DG/RMD calculations, peptide and amide-sulfur hydrogen bonds corresponding to slowly exchanging amide resonances were explicitly restrained (Qian & Weiss, 1992). Additional hydrogen bonds involving side-chain functionalities (e.g., Thr β-OH and amide groups of Asn and Gln; see below) may be inferred from the refined ensemble but do not correspond to slowly exchanging resonances in D_2O.

Side-Chain Environments. Refined DG/RMD models demonstrate that the TFIIS domain, like the classical Zn finger (Lee et al., 1989), folds as a globular minidomain with an inside and an outside. The positions of several side chains are well-defined (side-chain RMSD <1 Å; Figure 11A); these include residues 7, 9–13, 15, 19–23, 38–41, 43–45, and 47. Local environments of representative such side chains, shown as a stereo ensemble in Figure 14, are discussed in turn below.

An overview of side-chain relationships is provided by analysis of calculated solvent exposure, defined in Figure 11B as percent solvent accessibility. Predominantly exposed side chains (percent solvent accessibility greater than 60%) are either charged (Asp8, Lys14, Lys16, Lys17, Lys18, Arg28, Asp31, Glu32, Arg46, and Lys48) or near the chain termini (Thr5 and Cys50). Predominantly internal side chains (defined as percent accessibility less than 25%) include three of the coordinating cysteines (Cys12, Cys15, and Cys40), nonpolar and aromatic side chains (Phe10, Cys20, Tyr22, Phe37, Val38, and Trp47), selected polar side chains (Thr7, Thr21, Thr35, Thr36, Asn41, and Asn45), and one charged residue (Glu42). The Glu-42 carboxylate is solvent-exposed on one face and near the ϵ-NH_3^+ of Lys17; its methylene chain is engaged in long-range van der Waal contacts (see below). Similarly, the Thr γ-CH_3 groups contribute to nonpolar packing whereas their β-OH functions are either solvent-accessible or well-positioned to participate in hydrogen bonds (e.g., the β-OH of Thr21 is near the H_N protons of residues 5 and 6 and in some structures also near Asn41 δ-NH_2; see Table V). The overall pattern of *buried nonpolar groups* and *exposed polar or charged groups* leads to a favorable empirical solvation free energy (−29 ± 1 kcal/mol) as

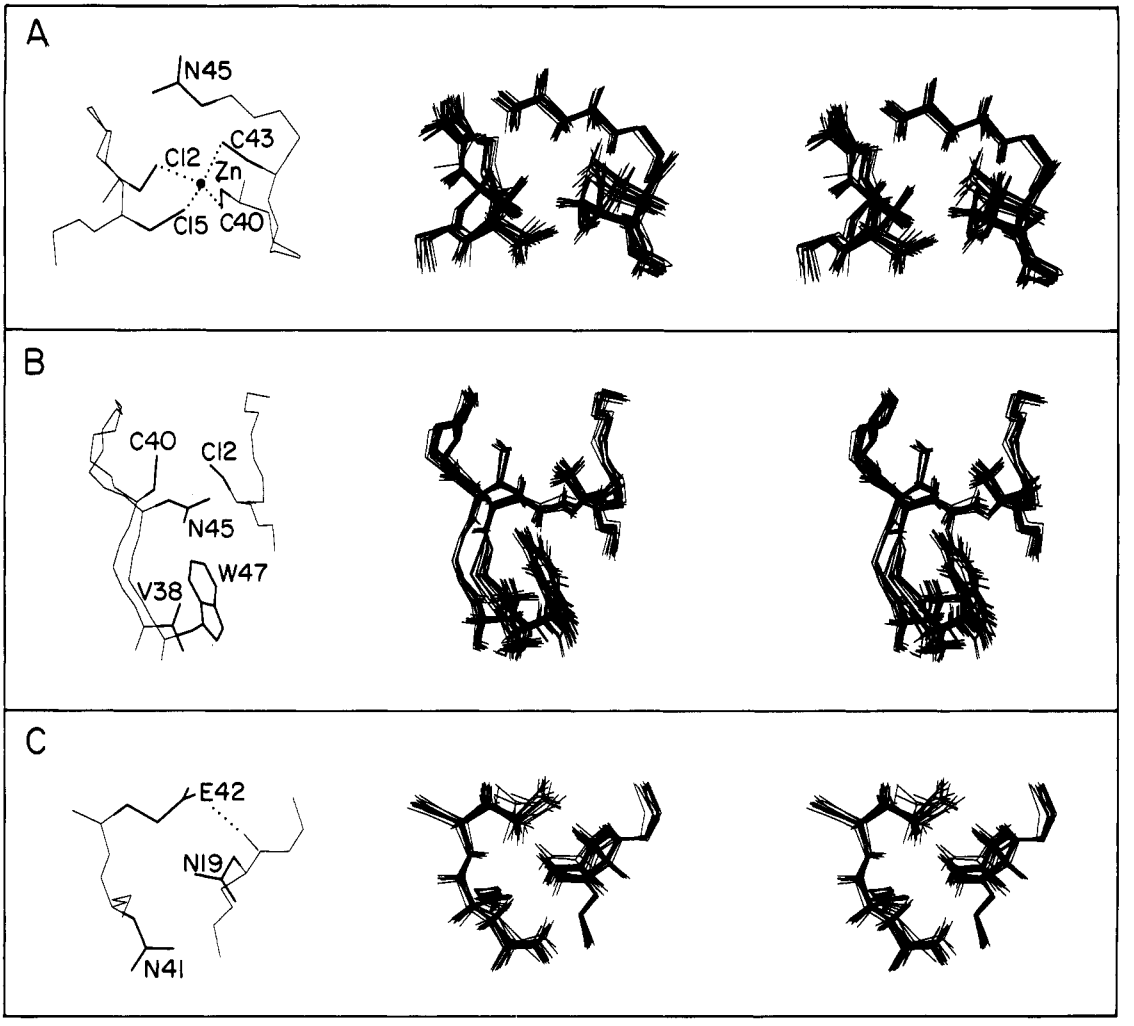

FIGURE 14: Local environments of well-ordered side chains. In each panel, one representative model is labeled at the left; the corresponding DG/RMD ensemble is shown in stereo at the right. (A) Zn coordination site. Thiolate-Zn bonds are shown as dotted lines. (B) Packing of N45 and W47 in β-strand 3 near V38 in β-strand 2. W47 also participates in an aromatic–aromatic interaction with F10 (not shown). (C) Contact between β-strands 1 and 2. Putative long-range hydrogen bond between the E42 carboxylate and main chain H_N of N19 is shown in the dotted line.

Table VI: Amide-Sulfur Hydrogen Bonds[a]

donor		acceptor		distance (Å)[b]	angle [Δ(deg)][c]
			Set A		
14	HN*	12	Sγ	2.5 ± 0.2	47 ± 7
15	HN*	12	Sγ	2.5 ± 0.2	13 ± 4
42	HN*	40	Sγ	2.4 ± 0.2	46 ± 5
43	HN*	40	Sγ	2.3 ± 0.1	19 ± 3
			Set B[d]		
(i) 17	HN	15	Sγ	4.3 ± 0.8	52 ± 9
45	HN	43	Sγ	4.0 ± 0.3	95 ± 7
(ii) 17	HN	15	Sγ	4.4 ± 0.8	51 ± 8
45	HN	43	Sγ	3.9 ± 0.3	93 ± 4

[a] In set A are given predicted amide-sulfur hydrogen bonds that correspond to slowly exchanging amide resonances in D_2O (asterisks). In set B are given additional amide-sulfur relationships which (although stereochemically permitted) do not correspond to slowly exchanging amide resonances. Two ensembles were calculated. In the first ensemble, minimum (Sγ, HN) internuclear distances in set A were set to be 2.2 Å; in the second ensemble, minimum (Sγ, HN) distances were set to be 2.2 Å for both set A and set B. The default minimum would otherwise be set to be the sum of hard-sphere radii (2.75 Å). [b] The values shown represent ensemble averages (±RMSD). [c] Δ(deg) is defined as deviation from linearity (±RMSD). [d] Subset B(i) was calculated from the first ensemble; subset B(ii) was calculated from the second ensemble. In both ensembles, neighboring Lys ϵ-NH_3^+ groups provide alternative electrostatic stabilization of the C15 and C43 Sγ atoms (see text).

calculated by the method of Eisenberg and McLachlan (1986).

The ensemble positions of side chains in or near the Zn-binding site are illustrated in Figure 14. The two Cys-X_2-Cys elements are shown in panel A. Although primarily internal, one cysteate side chain (Cys43) is partially exposed (mean solvent accessibility 35% ± 5). The geometry of Zn^{2+} coordination is only approximately tetrahedral. The six S_γ-Zn-S_γ bond angles fall into three groups (Table IV): ≈115° (Cys15-Zn-Cys40, Cys15-Zn-Cys43, and Cys40-Zn-Cys43), ≈107° (Cys12-Zn-Cys15 and Cys12-Zn-Cys43), and ≈97° (Cys12-Zn-Cys40). Such distortion from ideal tetrahedral geometry has previously been observed in Zn-binding sites both in solution (Omichinski et al., 1991) and in crystal structures (Eklund & Branden, 1983). The Cys12 and Cys40 thiolates each participate in bifurcating amide hydrogen bonds (Table VI and Figure 13; see above). Although not positioned to form hydrogen bonds, the S_γ atoms of Cys15 and Cys43 are near the ϵ-NH_3^+ groups of Lys17 and Lys14, respectively, presumably leading to favorable electrostatic interactions.

Adjacent to the Zn-binding site, the side chain of Asn45 is well-ordered and relates the Zn-binding site to the tertiary structure of the β-sheet. The Asn45 δ-NH_2 protons exhibit NOEs to Gly13-H_N, Lys14-H_N, and Lys14-$H_{\beta 1,\gamma 1}$ (Figure 14A) and also the aromatic ring of Trp47 (Figure 14B). The aromatic ring of Trp47 in turn contacts the Zn-binding site (the α proton of Cys12 and β protons of Cys40) and the side chain of Val38 in β-strand 2. The aromatic-methyl NOEs

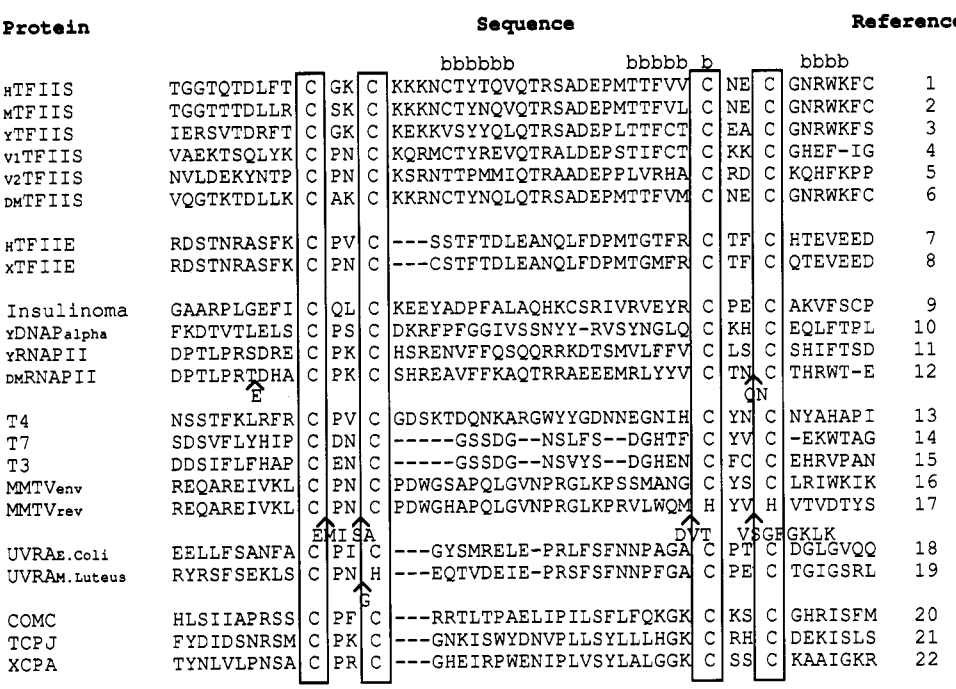

FIGURE 15: Alignment of putative Zn ribbon sequences obtained from databases. β-Sheet residues in TFIIS are designated by "b" at the top. Cysteines proposed to coordinate Zn^{2+} are aligned and boxed. Insertions are indicated by "^", and deletions are indicated by "–". (a) Sequence identification, reference, and database accession numbers are given below; superscripts identify source database ([SP]SWISS-PROT version 24.0 and [GB]GenBank release 74.0). (1) HTFIIS (human TFIIS): Yoo et al. (1991); TFS2_HUMAN[SP]. (2) MTFIIS (mouse TFIIS): Kanai et al. (1991); TFSM_MOUSE[SP]. (3) YTFIIS (yeast TFIIS): Hubert et al. (1983); Davies et al. (1990); TFS2_YEAST[SP]. (4) V1TFIIS (African swine fever virus TFIIS): Rodriguez et al. (1992); TFS2_ASFB[SP]. (5) V2TFIIS (stable subunit of vaccinia virus RNAP): Ahn et al. (1990); RPO5_VACCC[SP]. (6) DMTFIIS (*Drosophilia* TFIIS): Marshall et al. (1990); TFS2_DROME[SP]. (7) HTFIIE α-subunit of human TFIIE): Peterson et al. (1991); HUMTFIIEA_1[GB]. (8) XTFIIE (α-subunit of *Xenopus* laevis TFIIE): Ohkuma et al. (1992). (9) Insulinoma (putative neuroendocrine cell transformation protein, IA-1 gene product, *Homo sapiens*): Goto et al. (1992); HUM1A1X_1[GB]. (10) YDNAPa (*Saccharomyces cerevisiae* DNA polymerase α-subunit): Pizzagalli et al. (1988); DPOA_YEAST[SP]. (11) YRNAPII (*Saccharomyces cerevisiae* RNA polymerase II subunit): Woychik et al. (1991); RPB9_YEAST[SP]. (12) DMRNAPII (*Drosophilia melanogaster* RNA polymerase II 15-kDa subunit): Harrison et al. (1992); S88139_1[GB]. (13) T4 (bacteriophage T4 DNA primase): Nakanishi and Alberts (1985); Macdonald and Mosig (1984); DNAP_BPT4[SP]. (14) T7 (bacteriophage T7 DNA primase): Dunn and Studier (1983); Dunn and Studier (1981); PRIM_BPT7[SP]. (15) T3 (bacteriophage T3 DNA primase): Beck et al. (1989); PRIM_BPT3[SP]. (16) MMTVenv (mouse mammary tumor virus proviral envelop region): Moore et al. (1987); MMTENVGR_1[GB]. (17) MMTVrev (mouse mammary tumor virus reverse transcriptase): Moore et al. (1987); POL_MMTVB[SP]. (18) UVRAE.Coli (DNA-binding subunit of *E. coli* DNA-repair protein): Husain et al. (1986); UVRA_ECOLI[SP]. (19) UVRAM.Luteus (DNA-binding subunit of *Micrococcus luteus* DNA-repair protein): Shiota and Nakayama (1984); UVRA_MICLU[SP]. (20) COMC (late competence protein in *Bacillus subtilis*); Mohan et al. (1989); COMC_BACSU[SP]. (21) TCPJ (TCP pilin leader protease in *Vibrio cholerae*); Kaufman et al. (1991): TCPJ_VIBCH[SP]. (22) XCPA (secretion protein in *Pseudomonas aeruginosa*); Nunn et al. (1990); Bally et al. (1991); XCPA_PSEAE[SP].

are stereospecific. Edge-to-face packing is observed between the aromatic rings of Trp47 and Phe10 (Burley & Petsko, 1988). Such packing orients the $N_\epsilon H$ indole proton of Trp47 toward the center of the Phe10 π orbital, providing a weakly polar interaction whose strength is estimated to be similar to that of an internal hydrogen bond (Burley & Petsko, 1988). Analogous edge-to-face packing is possible between Phe10 and Phe49 but presently is less well-defined due to degeneracy of the F49 aromatic ^1H resonances (Table I); in the future this limitation may be overcome using 3D ^{13}C-edited NOESY experiments (Clore & Gronenborn, 1991). Aromatic–aromatic interactions are also observed in the classical Zn finger (Jasanoff et al., 1992).

The side chains of Asn41 and Glu42 (central residues in the C-terminal Cys-X_2-Cys element) are well-ordered near Asn19 (Figure 14C). In the NOESY spectrum, contacts are observed between Asn19 δ-NH_2 and Glu42-$H_{\gamma 1,2}$, between Asn19 δ-NH_2 and one Asn41-H_β, between Asn19 δ-NH_2 and Glu42-$H_{\gamma 1,2}$, and between one Asn19-H_β and Glu42-$H_{\gamma 1,2}$. The β protons of Asn19, Asn41, and Glu42 are stereospecifically assigned whereas the Glu42 γ protons are degenerate (Table I). In the DG/RMD ensemble, the Glu42 ϵ-COO^- is near the ϵ-NH_3^+ of Lys17 at the protein surface and may, in addition, accept a hydrogen bond from Asn19-H_N (Figure 14C). In the majority of structures, Asn41 δ-NH_2 donates a hydrogen bond to the β-OH of Thr21 (not shown in Figure 14; see Table V). In addition, an amide $N_\epsilon H$ of Gln24 is well-positioned to donate a hydrogen bond to Thr36-βOH (across the β-sheet) whereas the Gln24 amide oxygen is positioned to receive a hydrogen bond from Tyr22 p-OH (within β-strand 1). Although the latter hydrogen bonds are speculative, their formation is suggested by the presence of multiple NOEs among the sidechains of Tyr22, Gln24, and Thr36.

(IV) Database Search and Homology Modeling

TFIIS homologues have been cloned and sequenced from several organisms, including mouse, *Drosophila melanogaster*, and *Saccharomyces cerevisiae*; in addition, a TFIIS-like sequence occurs in vaccinia virus as a subunit of the viral RNAP (Figure 15; references are provided in the figure caption). These sequences exhibit striking conservation, which may reflect either the specific function of TFIIS in RNAPII elongation or general requirements of Zn-dependent folding. As a first step toward defining a Zn ribbon sequence template, we undertook a more general search for analogous sequences in other genes. Our search protocol had two parts. First, the available databases (GenBank, SWISS-PROT, and PIR; see Materials and Methods) were searched for strict matches to the degenerate template Cys-X_2-Cys-X_{6-9}-h-X_{14}-Cys-X_2-Cys-X_3-h (in which h designates a hydrophobic residue in accord with the sequence homologues). Matching sequences were then used to select more broadly related sequences from the

same databases. In the second search, departures from the initial template in ligand identity (Cys or His) or spacing (Cys-X_2-Cys or Cys-X_4-Cys) were tolerated as described (Higgins & Sharp, 1988, 1989). As shown in Figure 15, analogous sequences are found in a variety of genes involved in RNA or DNA transactions. These include basal initiation factor TFIIE, subunits of RNAPII, DNA primases of bacterial phages T3, T4, and T7, and bacterial DNA-repair proteins. The latter are of interest in light of a recently discovered coupling between transcription and DNA repair [for review, see Hanawalt and Mellon (1993)]. The classification of the above sequences as Zn-binding sites is only presumptive. The Cys_4 domain of T7 DNA primase (encoded by gene 5; Dunn & Studier, 1981, 1983) has recently been shown to bind a single Zn^{2+}; its CD spectrum is essentially identical to that of the TFIIS domain (L. Mendelman and C. A. Richardson, manuscript in preparation).

Formation of analogous β-sheet domains by the above sequences would imply that the Zn ribbon motif can accommodate a variety of substitutions, insertions, and deletions as required for divergent functions. To test the reasonableness of this proposal, a three-dimensional model of the TFIIE Cys_4 site was built using DG-based homology modeling (Havel, 1990). Its ligand spacing (Cys-X_2-Cys-$\mathbf{X_{21}}$-Cys-X_2-Cys) differs from that of TFIIS (Cys-X_2-Cys-$\mathbf{X_{24}}$-Cys-X_2-Cys). Optimal sequence alignment by the procedure of Higgins & Sharp (1988, 1989) suggested that a three-residue gap may be placed C-terminal to the second cysteine (Cys-X_2-Cys-**gap**-X_{21}-Cys-X_2-Cys; Figure 15). Following such alignment, the TFIIE and TFIIS sequences share 24% sequence identity and 39% sequence similarity. The proposed location of the gap has a plausible structural basis: the additional three residues in TFIIS (Lys16-Lys17-Lys18) form a loop between the N-terminal β-turn and the first β-strand (Figures 1B and 15). Neural-net analysis of the secondary structure (Hulley & Karplus, 1989) is compatible with β-sheet formation but also predicts a high probability of helix in the center of the sequence as previously noted (Peterson et al., 1991).

Distance-geometry models of the TFIIE domain were calculated based on 296 distance restraints and 52 dihedral restraints imposed by analogy to the TFIIS structure. Restraints include 42 (i,i+2) distances between α carbons and 215 long-range intercarbon distances (±15% of mean distances in the TFIIS ensemble). Long-range restraints were restricted to sites of amino acid identity (TFIIE residues Phe10, Cys12, Cys15, Thr21, Thr23, Pro32, Met33, Thr34, Cys40, and Cys43; using an analogous peptide numbering scheme) or chemical similarity (TFIIE residues Phe22, Leu25, Gln26, and Val47). The following additional restraints were imposed: peptide hydrogen-bond restraints in the β-sheet (11), amide-sulfur hydrogen bonds in the metal-binding site (4), and distance restraints related to tetrahedral Zn^{2+} coordination. Of 30 DG/SA calculations initiated, four convergent models were obtained. These exhibited an α-carbon RMSD of 2.0 Å (excluding the central loop and N-terminal residues) with negligible restraint violations. The models were subjected to restrained molecular dynamics and energy minimization (Clore et al., 1985). The individual components of the empirical energy function were similar in each case to those of the TFIIS ensemble (Table II).

A ribbon model of the TFIIE domain is shown in Figure 16 (model A). Alignment of the gap C-terminal to Cys15 is well-tolerated: foreshortening of the TFIIS loop ("gap" residues 16–19) enables formation of an analogous β-sheet and metal-binding site. The twist of β-strain 1 differs, however, from that of TFIIS as required to accommodate the deletion.

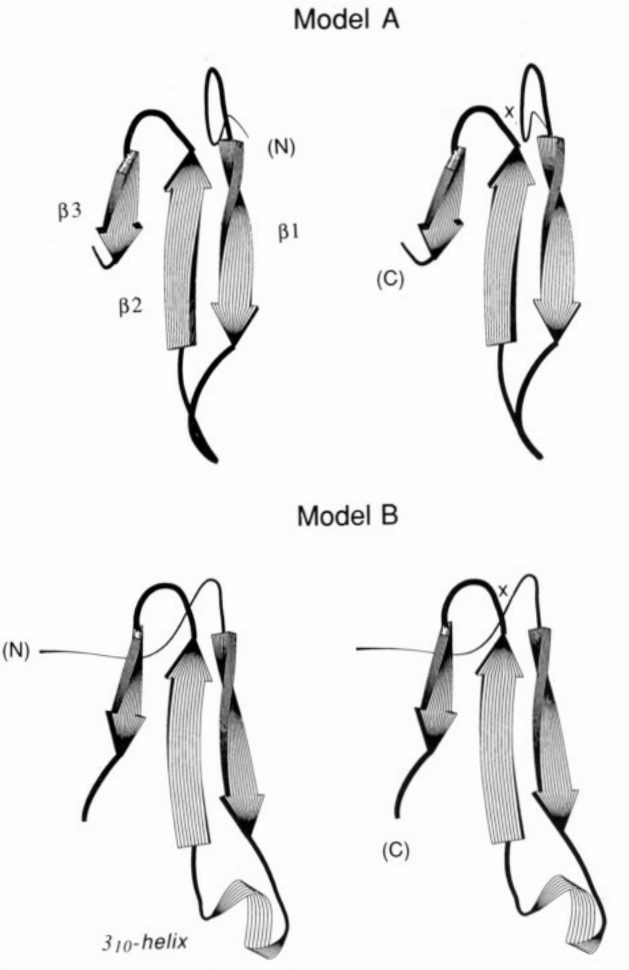

FIGURE 16: Ribbon models A (upper panel) and B (lower panel) of TFIIE Cys_4 domain as predicted by homology modeling. Strands β1–β3 are labeled in the upper left-hand panel. Model A, based strictly on constraints analogous to TFIIS, contains a disordered loop between strands β1 and β2. In model B, additional 3_{10}-helix related constraints were introduced in this region in accord with secondary structure predictions (Peterson et al., 1991). The position of Zn is indicated in the two models by an "X".

Detailed inspection of side-chain environments reveals no obvious biophysical violations (e.g., internal polar or charged residues). In fact, the empirical solvation free energy of the TFIIE models (–28 ± 1 kcal/mol; Eisenberg & McLachlan, 1986) is similar to that of the TFIIS ensemble, suggesting comparable desolvation of nonpolar surfaces. In model A the connection between the first and second β-strands is disordered as in TFIIS. In model B a short 3_{10}-helix has been introduced in accord with secondary-structure predictions (Peterson et al., 1991) and is well-tolerated. (Inclusion of α-helical hydrogen bonds are less readily accommodated; this distinction is likely to reflect the limitations of homology modeling and is not meant as a structural prediction.) Although such model building is plausible, extension of the Zn ribbon motif to TFIIE and analogous Cys_4 sites will require direct experimental demonstration.

DISCUSSION

Transcription requires an intricate choreography of protein–protein and protein–nucleic acid interactions, which are not understood in detail. Overall steps include promoter location, formation of a multiprotein initiation complex, "melting in", abortive initiation, elongation, and termination (Yager & von Hippel, 1991). A conceptual framework for analysis of elongation is provided by the "transcription bubble" para-

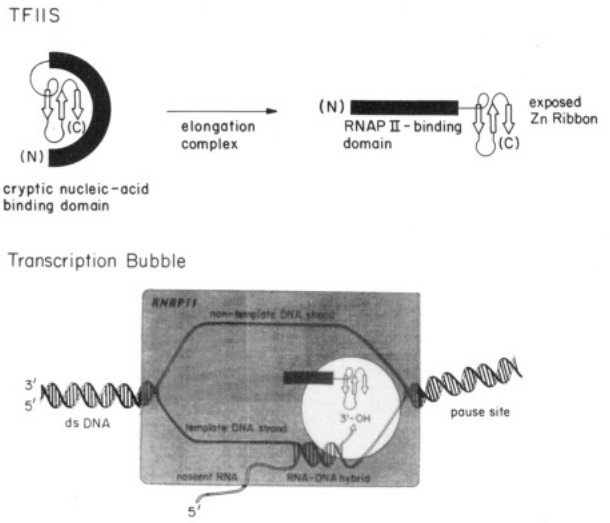

FIGURE 17: (A) Domain model of TFIIS based on biochemical studies of deletion mutants (Agarwal et al., 1991). The isolated N-terminal region (residues 1–229) binds to RNAP-II in an elongation complex whereas the isolated C-terminal domain (residues 230–280; the Zn ribbon) binds nucleic acids. The intact protein does not exhibit nucleic acid binding activity, motivating the hypothesis that the nucleic acid binding surface of the Zn ribbon is inaccessible (cryptic) in the structure of the intact protein (as illustrated in upper left). A conformational change is proposed on binding to an arrested elongation complex (upper right), permitting interaction between Zn ribbon and nucleic acid as schematically indicated in panel B (the transcription bubble paradigm).

digm: a complex of a template and nontemplate single-stranded DNA, the nascent RNA transcript, RNA polymerase, and associated factors (Figure 17B). This assembly moves processively along a gene in step with nascent RNA synthesis. Elongation must involve dynamic interactions among proteins and a variety of nucleic acids, including double-stranded DNA (upstream and downstream of the elongation bubble), single-stranded DNA (melted template and nontemplate strands), DNA–RNA heteroduplex (the RNAP catalytic site), and single-stranded RNA (the nascent message).

Elongation serves as a transcriptional control point in nuclear proto-oncogenes (Wright & Bishop, 1989) and viral genes, including HIV-1 (Hay et al., 1982; Hay & Aloni, 1984; Kao et al., 1987). Two eukaryotic elongation factors have been identified, TFIIS (also designated S-II; Natori, 1982) and TFIIF (also designated RAP30/74; Burton et al., 1988). These proteins are not homologous in sequence and have distinguishable activities. TFIIS, the object of the present structural study, binds to the largest subunit of RNAP-II in a paused elongation complex (Rappaport et al., 1988; Archambault et al., 1992). Such binding facilitates elongation through intrinsic DNA pause and termination sites (Reines et al., 1989; Siva-Raman et al., 1990; Agarwal et al., 1991) and past blocks imposed by sequence-specific DNA-binding proteins (Reines & Mote, 1993). Unlike TFIIS, TFIIF increases the overall rate of RNAPII elongation without specific effect at transcriptional pause sites (Bengal et al., 1991). Whereas TFIIS selectively interacts with the paused elongation complex, TFIIF can interact with RNAP-II in a preinitiation complex and is required for initiation of transcription (Burton et al., 1988; Fores et al., 1988).

Binding of TFIIS to a paused elongation complex activates a Mg^{2+}-dependent ribonuclease activity, resulting in cleavage of the nascent transcript (Reines & Mote, 1993; Izban & Luse, 1992). Purified TFIIS is not itself a ribonuclease. Analogous stimulation of a Mg^{2+}-dependent 3′→5′ ribonuclease is observed on binding of *E. coli* elongation factors (GreA and B) to cognate elongation complexes (Borukhov et al., 1992). Factor-stimulated RNA cleavage is distinct from pyrophosphorolysis (the back-reaction of RNAP-catalyzed formation of a 3′-5′ phosphodiester RNA linkage). Although otherwise unrelated in mechanism, both the TFIIS-stimulated ribonuclease activity and pyrophosphorolysis exhibit coupling between strand cleavage and retrograde movement of RNAP-II along the DNA template. Such coupling maintains correct alignment of the ternary complex (DNA template, foreshortened RNA strand, and RNAP-II active site), thus permitting the forward reaction to be resumed in register. TFIIS-coupled retrograde movement of RNAP-II is proposed as a necessary step in transcriptional read-through (Reines & Mote, 1993; Izban & Luse, 1992; Kassavetis & Geiduschek, 1993). TFIIF binding is not associated with nascent transcript cleavage.

How and where does TFIIS bind in an arrested elongation complex? Insight has been obtained from biochemical studies of mutants containing N-terminal, C-terminal, or central deletions (Agarwal et al., 1991). Such studies demonstrated the presence of at least two functional domains (Figure 17A).

(*i*) *RNAP-II-Binding Domain.* Interaction between TFIIS and RNAP-II, as assayed by the glycerol gradient technique of Reinberg and Roeder (1987), requires the presence of residues 100–229 (Agarwal et al., 1991). Deletion of this region leads to loss of activity in transcriptional elongation assays. Conversely, in the absence of the nucleic acid binding domain (below), residues 1–229 inhibit elongation. The sequence boundaries of the putative RNAP-II-binding domain and its three-dimensional structure have not been determined.

(*ii*) *Nucleic Acid Binding Domain.* Whereas intact TFIIS exhibits no detectable nucleic acid binding, such activity is readily observed by gel retardation assay on deletion of N-terminal sequences 1–174 or central sequences 103–171 (Agarwal et al., 1991). Nucleic acid binding is observed at high ionic strength ([NaCl] 250 mM) to ssDNA, dsDNA, ssRNA, and RNA–DNA heteroduplex; i.e., the range of nucleic acids found in the transcription bubble (Figure 17B). Such nucleic acid binding is retained by a C-terminal fragment containing a novel Cys_4 sequence (Figure 1A; residues 231–250). Its sequence specificity and mechanism of nucleic acid binding are not understood. In the intact protein, point mutations in the Zn ribbon have been identified that reduce antitermination activity (C.-J. Jeon and K. Agarwal, manuscript in preparation). The isolated nucleic acid binding domain does not bind to RNAP-II.

The above experiments demonstrate that protein sequences individually required for binding RNAP-II or nucleic acids are jointly required for biological activity (Agarwal et al., 1991). Because intact TFIIS exhibits stoichiometric Zn^{2+} binding, we propose that the Zn ribbon is folded in the native protein to provide a cryptic nucleic acid binding surface (Figure 17A). This surface is presumably exposed on binding of TFIIS to an elongation complex (Figure 17B) as previously proposed (Agarwal et al., 1991). We speculate that the Zn ribbon binds at or near the region of the RNA–DNA heteroduplex as shown; i.e., near the site of TFIIS-stimulated RNA cleavage. Analogous unmasking of a cryptic nucleic acid binding domain by deletion of occluding N-terminal sequences has recently been described in the σ^{70} subunit of *E. coli* RNA polymerase (Dombroski et al., 1992). In each case assembly-dependent exposure of functional surfaces may provide an allosteric mechanism to restrict sites of protein–nucleic acid interaction to the appropriate initiation (σ^{70}) or elongation (TFIIS) complex. Such regulation would circumvent the problem of nonspecific nucleic acid binding.

CONCLUSIONS

The solution structure of the TFIIS nucleic acid binding domain has been determined using homonuclear and heteronuclear 3D NMR methods. The structure consists of a novel three-stranded antiparallel β-sheet, which we designate the *Zn ribbon* (Qian et al., 1993). There is no relationship between the structure of the Zn ribbon and previously characterized Zn modules, such as the classical Cys_2-His_2 Zn finger (Miller et al., 1985; Brown et al., 1985; Klug & Rhodes, 1987; Berg, 1988; Gibson et al., 1988; Paragga et al., 1988; Lee et al., 1989; Pavletich & Pabo, 1991), retroviral CC/HC nucleocapsid domain (Summers et al., 1990; South et al., 1991), glucocorticoid-related Cys_4-Cys_4 "Zn twist" (Hare et al., 1990; Schwabe et al., 1990; Luisi et al., 1991; Vallee et al., 1991), or Gal4 Cys_6 binuclear cluster (Marmorstein et al., 1992; Kraulis et al., 1992; Baleja et al., 1992). Putative Cys_4 Zn-binding sites analogous to that of TFIIS are predicted in a variety of proteins involved in nucleic acid transactions, including subunits of RNAP-II and eukaryotic transcriptional initiation factor TFIIE. The present study of the TFIIS Zn ribbon thus defines a new structural motif and provides a foundation for future investigations of TFIIS–nucleic acid interactions. Such interactions are of interest in relation to its novel mechanism of antitermination by stimulation of nascent RNA cleavage.

ADDED IN PROOF

An analogy may be drawn between the TFIIS Zn ribbon and the structure of the *B. subtilis* major cold-shock protein (Schindelin et al., 1993; Schnuchel et al., 1993). The latter contains five β-strands, which form two subdomains (β1–β3 and β4–β5); this motif is designated a "universal nucleic acid binding domain" due to its homology to eukaryotic Y-box transcription factors. Although the cold-shock protein does not contain Zn and is unrelated to TFIIS in sequence, the putative nucleic acid binding surface of its β1–β3 subdomain contains multiple aromatic and basic amino acids. TFIIS strands β1–β3 define an analogous surface. The TFIIS structure is unrelated to the recently described structure of the GATA Cys_4 DNA-binding motif (Omichinski et al., 1993).

ACKNOWLEDGMENT

NMR spectra were obtained at Harvard Medical School (NIH 1 S10 RR04862-01). We thank J.-M. Chandonia, A. Sali, and M. Karplus for neural-net calculations; T. Havel (Harvard Medical School) for the program DG-II; Q. X. Hua and J. P. Lee for assistance with 3D NMR measurements and processing; L. Mendelman and C. A. Richardson for communication of results prior to publication; and L. J. Neuringer (deceased), K. Struhl, G. Wagner, and I. Wool for helpful discussion.

SUPPLEMENTARY MATERIAL AVAILABLE

Figures (3) showing slowly exchanging amide resonances and a table of NOE and dihedral restraints used for DG/SA calculation (13 pages). Ordering information is given on any current masthead page.

REFERENCES

Adler, M., Carter, P., Lazarus, R. A., & Wagner, G. (1993) *Biochemistry 32*, 282–289.
Agarwal, K., Baek, K.-H, Jeon, C.-J., Miyamoto, K., Ueno, A., & Yoon, H.-S. (1991) *Biochemistry 30*, 7842–7851.
Ahn, B.-Y., Gershon, P. D., Jones, E. V., & Moss, B. (1990) *Mol. Cell. Biol. 10*, 5433–5441.
Akke, M., Drakenberg, T., & Chazin, W. J. (1992) *Biochemistry 31*, 1011–1020.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403–410.
Archambault, J., Lacroute, F., Rluet, A., & Friesen, J. D. (1992) *Mol. Cell. Biol. 12*, 4142–4152.
Baek, K.-H., Sato, K. S., Ito, R., & Agawal, K. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7623–7627.
Bally, M., Ball, G., Badere, A., & Lazdunksi, A. (1991) *J. Bacteriol. 173*, 479–486.
Beck, P. J., Gonzalez, S., Ward, C. L., & Molineux, I. J. (1989) *J. Mol. Biol. 210*, 687–701.
Bengal, E., Flores, O., Krauskopf, A., Reinberg, D., & Aloni, A. (1991) *Mol. Cell. Biol. 11*, 1195–1206.
Bentley, D. L., & Groudine, M. (1986) *Nature 321*, 702–706.
Bentley, D. L., & Groudine, M. (1988) *Cell 53*, 245–256.
Berg, J. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 99–487.
Borukhov, S., Polyakov, A., Nikiforov, V., & Goldfarb, A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 8899–8902.
Bowie, J. U., Lüthy, R., & Eisenberg, D. (1991) *Science 253*, 164–170.
Burley, S. K., & Petsko, G. A. (1988) *Adv. Protein Chem. 39*, 125–189.
Burton, Z. F., Sopta, M., Ortolan, L. G., & Greenblatt, J. (1988) *Mol. Cell. Biol. 8*, 1602–1613.
Chinsky, J. M., Maa, M.-C., Ramamurthy, V., & Kellems, R. E. (1989) *J. Biol. Chem. 264*, 14561–14565.
Clore, G. M., & Gronenborn, A. M. (1991) *Prog. Nucl. Magn. Reson. Spectrosc. 23*, 43–92.
Davies, C. J., Tzrgovcich, J., & Hutchison, C. A., III (1990) *Nature 345*, 298.
Dombroski, A. J., Waiter, W. A., Record, M. T., Jr., Siegele, D. A., & Gross, C. A. (1992) *Cell 70*, 501–512.
Dunn, J. J., & Studier, F. W. (1981) *J. Mol. Biol. 148*, 303–330.
Dunn, J. J., & Studier, F. W. (1983) *J. Mol. Biol. 166*, 477–535.
Eisenberg, D., & McLachlan, A. D. (1986) *Nature 319*, 199–203.
Eklund, H., & Branden, C.-I. (1983) in *Zinc Enzymes* (Spiro, T. G., Ed.) John Wiley, New York.
Fores, O., Maldonado, E., Burton, Z., Greenblatt, J., & Reinberg, D. (1988) *J. Biol. Chem. 263*, 10812–10816.
Goto, Y., De Silva, M. G., Toscani, A., Prabhakar, B. S., Notkins, A. L., & Lan, M. S. (1992) *J. Biol. Chem. 267*, 15252–15257.
Hanawalt, P., & Mellon, I. (1993) *Curr. Biol. 3*, 67–69.
Harrison, D. A., Mortin, M. A., & Corces, V. G. (1992) *Mol. Cell. Biol. 12*, 928–935.
Havel, T. F. (1990) *Biopolymers 29*, 1565–1585.
Havel, T. F. (1991) *Prog. Biophys. Mol. Biol. 56*, 43–78.
Hay, N., & Aloni, Y. (1984) *Nucleic Acids Res. 12*, 1401–1413.
Hay, N., Skolnik-David, H., & Aloni, Y. (1982) *Cell 29*, 183–193.
Higgins, D. G., & Sharp, P. M. (1988) *Gene 73*, 237–244.
Higgins, D. G., & Sharp, P. M. (1989) *CABIOS 5*, 151–153.
Hirashima, S., Hirai, H., Nakanishi, Y., & Natori, S. (1988) *J. Biol. Chem. 263*, 3858–3863.
Horikoshi, M., Sekimizu, K., & Natori, S. (1984) *J. Biol. Chem. 259*, 608–611.
Hubert, J. C., Guyonvarah, A., Kammerer, B., Exinger, F., Liljolund, P., & Lacroute, F. (1983) *EMBO J. 2*, 2071–2073.
Husain, I., Van Houten, B., Thomas, D. C., & Sancar, A. (1986) *J. Biol. Chem. 261*, 4895–4901.
Izban, M. G., & Luse, D. S. (1992) *Genes Dev. 6*, 1342–1356.
Jasanoff, A., Kochoyan, M., Lee, J. P., & M. A. Weiss (1992) *J. Mol. Biol. 225*, 1035–1047.
Kanai, A., Kuzuhara, T., Sekimizu, K., & Natori, S. (1991) *J. Biochem. 109*, 674–677.
Kao, S.-Y., Calman, A. F., Luciw, P. A., & Peterlin, B. M. (1987) *Nature 330*, 489–493.
Kassavetis, G. A., & Geiduschek, E. P. (1993) *Science 259*, 944–945.
Kaufman, M. R., Seyer, J. M., & Taylor, R. K. (1991) *Genes Dev. 5*, 1834–1846.

Kerppola, T. K., & Kane, C. M. (1988) *Mol. Cell. Biol. 8*, 4389–4394.
Kochoyan, M., Keutmann, H. T., & Weiss, M. A. (1991) *Biochemistry 30*, 7063–7072.
Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., & Wright, P. E. (1989) *Science 245*, 635–637.
Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., & Sigler, B. P. (1991) *Nature 352*, 497–505.
Macdonald, P. M., & Mosig, G. (1984) *EMBO J. 3*, 2863–2871.
Marion, D., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989) *J. Mag. Reson. 85*, 393–399.
Marmorstein, R., Carey, M., Ptshne, M., & Harrison, S. C. (1992) *Nature 356*, 408–414.
Marshall, J. K., Guo, H., & Price, D. H. (1990) *Nucleic Acids Res. 18*, 6293–6298.
Mather, E. L., Nelson, K. J., Haimovich, J., & Perry, R. P. (1984) *Cell 36*, 329–338.
Messerle, B., Wider, G., Otting, G., Weber, C., & Wuthrich, K. (1989) *J. Mag. Reson. 85*, 608.
Mohan, S., Aghion, J., Guillen, N., & Dubnau, D. (1989) *J. Bacteriol. 171*, 6043–6051.
Moore, R., Dixon, M., Smith, R., Peters, L., & Dickson, C. (1987) *J. Virol. 61*, 480–490.
Natori, S. (1982) *Mol. Cell. Biochem. 46*, 173–187.
Natori, S., Takeuchi, K., & Mizuno, D. (1973) *J. Biochem. (Tokyo) 74*, 1177–1182.
Nunn, D., Bergman, S., & Lory, S. (1990) *J. Bacteriol. 172*, 2911–2919.
Ohkuma, Y., Sumimoto, H., Hiffman, A., Shimasaki, S., Horikoshi, M., & Roeder, R. G. (1991) *Nature 354*, 398–401.
Ohkuma, Y., Hashimoto, S., Roeder, R. G., & Horikoshi, M. (1992) *Nucleic Acids Res. 20*, 4363.
Omichinski, J. G., Clore, G. M., Apella, E., Sakaguchi, K., & Gronenborn, A. M. (1990) *Biochemistry 29*, 9324–9334.
Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., & Gronenborn, A. M. (1993) *Science 261*, 438–446.
Oschkinat, H., Cieslar, C., Gronenborn, A. M., & Clore, G. M. (1989a) *J. Magn. Reson. 81*, 212.
Oschkinat, H., Cieslar, C., Holak, T. A., Clore, A. M., & Gronenborn, A. M. (1989b) *J. Magn. Reson. 83*, 450.
Pabo, C. O., & Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053–1095.
Parkhurst, S. M., Harrison, D. A., Remington, M. P., Spana, C., Kelley, R. L., Coyne, R. S., & Corces, V. G. (1988) *Genes Dev. 2*, 1205–1215.
Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., & Klevit, R. E. (1988) *Science 241*, 1489–1492.
Peterson, M. G., Inostroza, J., Maxon, M. E., Flores, O., Admon, A., Reinberg, D., & Tjian, R. (1991) *Nature 354*, 369–373.
Pizzagalli, A., Valsasnini, P., Blevani, P., & Lucchini, G. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3772–3776.
Qian, X., & Weiss, M. A. (1992) *Biochemistry 31*, 7463–7476.
Qian, X., Jeon, C. J., Yoon, H.-S., Agarwal, K., & Weiss, M. A. (1993) *Nature (London)* (in press).
Rappaport, J., Reinberg, D., Zandomeni, R., & Weinmann, R. (1987) *J. Biol. Chem. 262*, 5227–5232.
Rappaport, J., Cho, K., Saltzman, A., Prenger, J., Golomb, M., & Weinmann, R. (1988) *Mol. Cell. Biol. 8*, 3136–3142.
Reinberg, D., & Roeder, R. G. (1987) *J. Biol. Chem. 262*, 3331–3337.
Reines, D., & Mote, J., Jr. (1993) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
Reines, D., Chamberlin, M. J., & Kane, C. M. (1989) *J. Biol. Chem. 264*, 10799–10809.
Resnekov, O., & Aloni, Y. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 12–16.
Rodriguez, J. M., Salus, M. L., & Vinuela, E. (1992) *Virology 186*, 40–52.
Rougvie, A. E., & Lis, J. T. (1988) *Cell 54*, 795–804.
Sawadogo, M., Sentenac, A., & Fromageot, P. (1980) *J. Biol. Chem. 255*, 12–15.
Schindelin, H., Marahiel, M. A., & Heinemann, U. (1993) *Nature 364*, 164–168.
Schnuchel, A., Wiltscheck, R., Czisch, M., Herrier, M., Willimsky, G., Graumann, P., Marahiel, M. A., & Holak, T. A. (1993) *Nature 364*, 169–171.
Schwabe, J. W. R., Neuhaus, D., & Rhodes, D. (1990) *Nature 348*, 458–461.
Sekimizu, K., Nakanishi, Y., Mizuno, D., & Natori, S. (1979) *Biochemistry 18*, 1582–1587.
Shiota, S., & Nakayama, H. (1989) *Mol. Gen. Genet. 217*, 332–340.
Siva-Raman, L., Reines, D., & Kane, C. M. (1990) *J. Biol. Chem. 265*, 14554–14560.
Sluder, A. E., Greenleaf, A. L., & Price, D. H. (1989) *J. Biol. Chem. 264*, 8963–8969.
South, T. L., Blake, P. R., Hare, D. R., & Summers, M. F. (1991) *Biochemistry 30*, 6342–6349.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982), *J. Magn. Reson. 48*, 286–292.
Summers, M. F., South, T. L., Kim, B., & Hare, D. R. (1990) *Biochemistry 29*, 329.
Vallee, B. L., Coleman, J. E., Auld, D. S. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 999–1003.
Woychik, N. A., Lane W. S., & Young, R. A. (1991) *J. Biol. Chem. 266*, 19053–19055.
Wright, S., & Bishop, J. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 505–509.
Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Yager, T. D., & von Hippel, P. H. (1991) *Biochemistry 30*, 1097–1118.
Yanofsky, C. (1988) *J. Biol. Chem. 263*, 609–612.
Yoo, O.-J., Yoon, H.-S., Baek, K.-H., Jeon, C.-J., Miyamoto, K., Ueno, A., & Agarwal, K. (1991) *Nucleic Acids Res. 19*, 1073–1079.