Structural similarity between TAFs and the heterotetrameric core of the histone octamer

Xiaoling Xie, Tetsuro Kokubo, Steven L. Cohen, Urooj A. Mirza, Alexander Hoffmann, Brian T. Chait, Robert G. Roeder, Yoshihiro Nakatani & Stephen K. Burley

Reprinted from Nature, Vol. 380, March 28, 1996

Structural similarity between TAFs and the heterotetrameric core of the histone octamer

Xiaoling Xie*†, Tetsuro Kokubo‡#, Steven L. Cohen§, Urooj A. Mirza§, Alexander Hoffmann‖#, Brian T. Chait§, Robert G. Roeder‖, Yoshihiro Nakatani‡ & Stephen K. Burley*†¶

Laboratories of * Molecular Biophysics, § Mass Spectrometry and Gaseous Ion Chemistry, and ‖ Biochemistry and Molecular Biology, and † Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA
‡ National Institutes of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

A complex of two TFIID TATA box-binding protein-associated factors (TAF_{II}s) is described at 2.0 Å resolution. The amino-terminal portions of $dTAF_{II}42$ and $dTAF_{II}62$ from *Drosophila* adopt the canonical histone fold, consisting of two short α-helices flanking a long central α-helix. Like histones H3 and H4, $dTAF_{II}42$ and $dTAF_{II}62$ form an intimate heterodimer by extensive hydrophobic contacts between the paired molecules. In solution and in the crystalline state, the $dTAF_{II}42/dTAF_{II}62$ complex exists as a heterotetramer, resembling the $(H3/H4)_2$ heterotetrameric core of the histone octamer, suggesting that TFIID contains a histone octamer-like substructure.

EUKARYOTIC transcription initiation and its regulation are best understood for genes transcribed by RNA polymerase II (Pol II) with the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIG/J, TFIIH and TFIII (reviewed in refs 1,2). In some cases, this process begins with recognition of the TATA element within the core promoter by the DNA-binding subunit of TFIID (TATA box-binding protein, TBP), forming a multiprotein–DNA complex that coordinates accretion of class II initiation factors and Pol II into a preinitiation complex (PIC) (reviewed in ref. 3). TFIIB is the next general factor to enter the PIC, creating a TFIIB–TFIID–DNA platform that is recognized by Pol II plus TFIIF. *In vivo*, Pol II transcription depends on TFIIE and TFIIH, and possibly TFIIA. Once PIC assembly is complete, and nucleoside triphosphates are present, strand separation occurs to give an open complex, the large subunit of Pol II is phosphorylated, and Pol II initiates transcription and is

¶To whom correspondence should be addressed.
Present address: Division of Gene Function in Animals, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-01, Japan (T.K.); Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (A.H.).

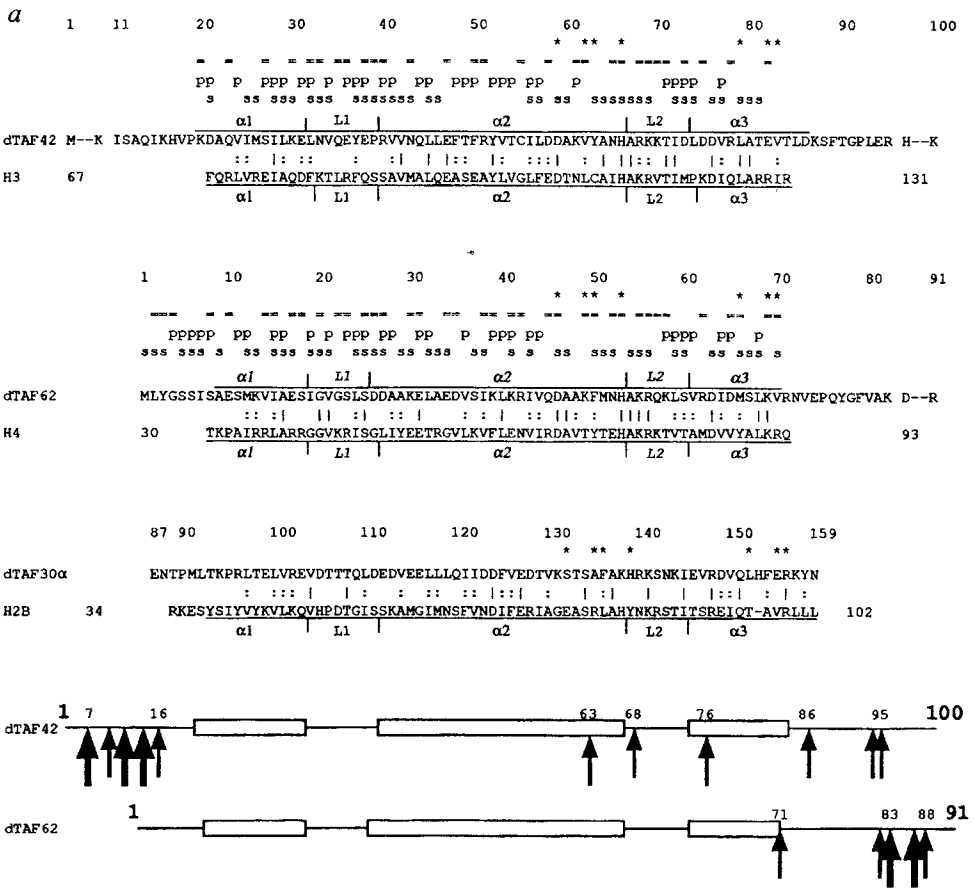


FIG. 1 *a*, Top, sequence alignments of histone-like portions of dTAF$_{II}$30α, dTAF$_{II}$42 and dTAF$_{II}$62 with chicken H2B, H3 and H4, respectively. Amino-acid identities are denoted by vertical lines and similarities with colons. The α-helical regions were assigned from X-ray structures[34]. Residues are labelled with 's', 'p' and an asterisk, indicating involvement in intramolecular, heterodimer, and heterotetramer contacts, respectively. For H2B, H4, dTAF$_{II}$30α and dTAF$_{II}$62, the asterisk denotes residues involved in contacts between distinct heterodimers. Solvent-accessible residues in the (dTAF$_{II}$42(17–86)/(dTAF$_{II}$62(1–70))$_2$ heterotetramer are denoted by short black bars. Bottom, proteolysis of the dTAF$_{II}$42(1–100)/dTAF$_{II}$62(1–91) complex. Endoproteases Asp-N, Glu-C, trypsin, subtilisin and chymotrypsin were used as previously described[29]. The polypeptide chains are represented schematically with large arrows denoting cleavage sites observed within minutes to hours, and small arrows denoting cleavage sites observed between 4 and 24 hours. *b*, Sequence comparisons of dTAF$_{II}$42(22–83), dTAF$_{II}$62(9–70), H3(68–130) and H4(31–92). The upper triangle gives the number of identical amino acids, and the lower triangle gives the number of chemically similar amino acids. Bold indicates comparisons between homologous proteins, and italics denote comparisons between heterodimerization partners. c, Pairwise comparisons of dTAF$_{II}$s and their histone homologues. The upper triangle gives r.m.s.ds between HSH2 portions, and the lower triangle gives the corresponding values for the HSH1 portions. *d*, Equilibrium analytical ultracentrifugation of dTAF$_{II}$42(11–95)/dTAF$_{II}$62(1–82). The fitted curve is superimposed on the observed absorption versus radius profile, and the calculated residuals are given in the top panel. The equilibrium dissociation constant (K_d) describing the transition from the (dTAF$_{II}$42(11–95)/dTAF$_{II}$62(1–82))$_2$ heterotetramer to the dTAF$_{II}$42(11–95)/dTAF$_{II}$62(1–82) heterodimer is about 10^{-6} M (goodness of fit, 0.00012), Similar K_d values were obtained for (H3/H4)$_2$ to 2(H3/H4) and its reverse reaction[35].

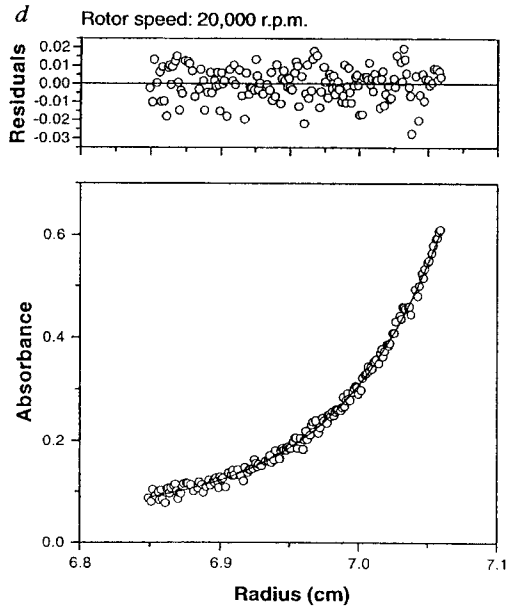
METHODS. *Drosophila* TAF$_{II}$42 residues 11–95 and *Drosophila* TAF$_{II}$62 residues 1–82 fused with GST were separately overexpressed as inclusion bodies in *E. coli*. Denatured dTAF$_{II}$42(11–95) and GST-dTAF$_{II}$62(1–82) were mixed in an equimolar ratio, co-renatured by stepwise dialysis, treated with protease to remove the GST tag, and purified to homogeneity by conventional methods. Mass spectrometry showed that the dTAF$_{II}$42(11–95)/dTAF$_{II}$62(1–82) complex used successfully for crystallization was neither modified nor further proteolysed during expression, renaturation and purification. The measured relative molecular mass values were: dTAF$_{II}$42(11–95), 9,917 ± 1 (predicted, 9,918); and dTAF$_{II}$62(1–82), 9,071 ± 1 (predicted, 9,072). Dynamic light scattering was performed with a Molecular Size Detector (Protein Solutions, Charlottesville, VA). Equilibrium analytical ultracentrifugation was performed using a Beckman model XL-A ultracentrifuge.

b

	dTAF42	dTAF62	H3	H4
dTAF42		*14*	**14**	12
dTAF62	*16*		13	**15**
H3	**18**	14		*12*
H4	17	**17**	*15*	

c

	dTAF42	dTAF62	H3	H4
dTAF42		0.7	0.9	0.7
dTAF62	1.4		1.0	0.7
H3	1.9	1.8		1.0
H4	1.4	1.7	1.9	



released from the promoter. During elongation *in vitro*, TFIID can remain bound to the core promoter and support rapid reinitiation of transcription (reviewed in ref. 4). An abbreviated PIC assembly mechanism has also been proposed, following discoveries of various Pol II holoenzymes (reviewed in ref. 5).

The role of TFIID in eukaryotic transcription has made it the focus of biochemical and genetic study since its discovery[6]. DNA binding by human TFIID was first demonstrated with the adenovirus major late promoter (AdMLP).[7] DNase I footprinting studies of the AdMLP and selected human promoters revealed

TABLE 1 Statistics of the crystallographic analysis

	Resolution (Å)	Reflections measured/unique	Completeness (%) overall/outer shell	R_{sym} (%) overall/outer shell	R_{iso}	Phasing power	R_c
(a) Data set							
MAD analysis (4 Se sites)							
λ1 (0.9873 Å)	18.0–2.4	34,011/7,697	91.0/94.7	5.9/8.0			
λ2 (0.9795 Å)	18.0–2.4	33,699/7,643	91.2/95.0	6.7/9.1			
λ3 (0.9792 Å)	18.0–2.4	33,786/7,623	91.1/94/7	7.6/9.9			
λ4 (0.9718 Å)	18.0–2.4	31,410/7,578	89.9/94.1	6.4/8.7			
		Overall MAD figure of merit, 0.66					
MIR analysis							
PIP (3 Pt and 1 I sites)	15.0–2.4	62,639/8,245	98.1/98.7	5.0/8.6	0.171	0.93	0.63
Hg (2 Hg sites)	15.0–2.4	37,298/8,111	96.5/99.9	5.2/6.9	0.178	0.49	0.78
K_2PtCl_6 (2 Pt sites)	15.0–2.4	29,231/7,808	93.3/85.9	3.9/6.8	0.127	0.55	0.74
K_2PtCl_4 (2 Pt sites)	15.0–2.4	40,498/7,838	93.7/93.8	6.0/10.1	0.191	0.65	0.72
		Overall MIR figure of merit, 0.51					
		Overall MAD and MIR combined figure of merit, 0.73					
Native data	18.0–1.8	156,415/19,413	99.8/99.6	5.3/17.3			

(b) Refinement statistics	Resolution (Å)	Completeness (%)	R-factor (overall/outer shell)	Free R-factor
Data with $\|F\| > 2\sigma(\|F\|)$	6.0–2.0	89.8	0.198/0.218	0.244
R.m.s. deviation	Bond lengths, 0.009 Å	Bond angles, 1.2°	Thermal parameters, 1.7 Å^2	

Crystals of the dTAF$_{II}$42(11–95)/dTAF$_{II}$62(1–82) complex containing a single mutation (Cys 55 → Ser in dTAF$_{II}$42) were obtained using hanging-drop vapour diffusion. The dTAF$_{II}$42(11–95) and dTAF$_{II}$62(1–82) constructs were also expressed in selenomethionine (Se–Met) substituted forms[50], and the Se–Met complex was purified as described above. A mercury adduct of the dTAF$_{II}$42(11–95)/dTAF$_{II}$62(1–82) complex was prepared by dialysing purified wild-type dTAF$_{II}$42(11–95)/dTAF$_{II}$(1–82) complex against $HgCl_2$. Oscillation photography at 100K and data reduction were performed as previously described[8]. Native data were collected using Beamline X12B, at the National Synchrotron Light Source, Brookhaven National Laboratory (BNL). Se–Met MAD data and MIR data from the mercury derivative were collected at BNL using Beamline X4A. The remaining heavy-atom derivative data were collected as previously described[11]. MAD[50] and MIR (MLPHARE; Z. Otwinowski) phase sets were calculated separately giving figure of merit (f.o.m.) values of 0.66 and 0.51, respectively. Phase combination yielded a final overall f.o.m. of 0.73 at 2.4 Å resolution, giving a high-quality experimental electron-density map revealing continuous electron density for almost the entire binary complex. Model building interspersed with positional refinement allowed an unambiguous trace and sequence assignment dTAF$_{II}$42(17–86) and dTAF$_{II}$62(1–70). The remaining 6 N-terminal and 9 C-terminal residues of dTAF$_{II}$42, and 12 C-terminal residues of dTAF$_{II}$62, were disordered and were omitted from the current model. The average *B*-factor for all protein atoms is 18 Å^2, the 125 water molecules included in the model have an average *B*-factor of 29 Å^2, and the 7 Zn^{2+} ions have an average *B*-factor of 28 Å^2. Atomic coordinates and structure factor amplitudes will be submitted to the Brookhaven Protein Data Bank. $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is the observed intensity, and $\langle I\rangle$ is the average intensity obtained from multiple observations of symmetry related reflections. Mean fractional isomorphous difference $= \Sigma||F_{PH}| - |F_P||/\Sigma|F_P|$, where $|F_P|$ is the protein structure factor amplitude, and $|F_{PH}|$ is the heavy-atom derivative structure factor amplitude. Phasing power = r.m.s.$(|F_H|/E)$, where $|F_H|$ is the heavy-atom structure factor amplitude and E is the residual lack of closure. R.m.s, bond lengths and r.m.s. bond angles are the respective root-mean-square deviations from ideal values. R.m.s. thermal parameter is the r.m.s.d. between the *B*-values of covalently bonded atomic pairs. Free *R*-factor was calculated with 10% of data omitted from the structure refinement. PIP, di-*m*-idobis(ethylenediamine)diplatinum(II) nitrate (Strem Chemicals).

sequence-specific interactions between TFIID and the TATA element that are primarily mediated by TBP (reviewed in ref. 8). In contrast, protection both upstream and downstream of the TATA box is largely sequence independent, displays a nucleosome-like pattern of DNase I hypersensitivity, differs between promoters (for example, AdMLP and the human heat-shock protein 70 (Hsp70) promoter protect residues −47 to +35 and −35 to −19, respectively), and is induced by some activators (reviewed in ref. 9). TATA box binding by TFIID or TBP precludes packaging of the core promoter with histone proteins (H2A, H2B, H3 and H4). Conversely, core promoter packaging by histone octamers prevents TFIID or TBP binding to the TATA element, effectively repressing transcription (reviewed in ref. 10).

Publication of the sequence of yeast TBP was followed rapidly by the sequences of homologous genes from various eukaryotes and an archaebacterium (amino-acid identities within the carboxy-terminal portion range between 38 and 100%; reviewed in ref. 11), and considerable efforts have been directed towards understanding the mechanisms by which TBP acts in Pol II transcription. Recombinant TBP alone can bind both general and regulatory factors and direct PIC assembly and basal transcription (reviewed in refs 8, 11). Activator-dependent transcription, however, requires TBP and the remaining subunits of TFIID, the TBP-associated factors (TAF$_{II}$s), and some non-TAF$_{II}$ coactivators (reviewed in ref. 9). Gene disruption studies of four yeast TAF$_{II}$s demonstrated that they are essential[12,13].

Affinity purification of TFIID allowed the identification of a set of phylogenetically conserved TAF$_{II}$s, denoted by their origins and relative molecular mass (reviewed in ref. 9). In both human and *Drosophila*, TAF$_{II}$s are tightly associated with TBP, providing binding sites for many different transcriptional activators and coactivators that modulate transcription initiation by Pol II by means of specific protein–protein interactions with TFIID (reviewed in ref. 9). Reconstitution of TFIID has been achieved *in vitro*[14], providing important insights into the roles of individual subunits as activator-specific targets that facilitate TFIID recruitment[15].

Primary structure analyses of some TAF$_{II}$s have indicated considerable sequence homology with non-linker histone proteins (Fig. 1)[16–19]. In *Drosophila*, dTAF$_{II}$42 (refs 16, 20) and dTAF$_{II}$62 (refs 16, 21) resemble H3 and H4, respectively, and correspond to human hTAF$_{II}$31 (ref. 17) and hTAF$_{II}$80 (refs 17, 21). Both *Drosophila* and human TFIID also contain TAF$_{II}$s (dTAF$_{II}$30α and hTAF$_{II}$20)[16,18,22,23] that are putative histone H2B homologues[18], but appear to lack histone H2A homologues. The TAF$_{II}$ nomenclatures of Roeder and Tjian (reviewed in ref. 9) have been adopted for human and *Drosophila* TFIID, respectively, the only exceptions being dTAF$_{II}$62 and dTAF$_{II}$42 which are denoted dTAF$_{II}$60 and dTAF$_{II}$40 by Tjian. We have previously documented a structural connection between eukaryotic transcription and DNA packaging. The co-crystal structure of the DNA-binding domain of the liver-specific transcription factor hepatocyte nuclear factor (HNF) 3-γ (ref. 24) resembles the structure of the linker histone H5 obtained without DNA[25].

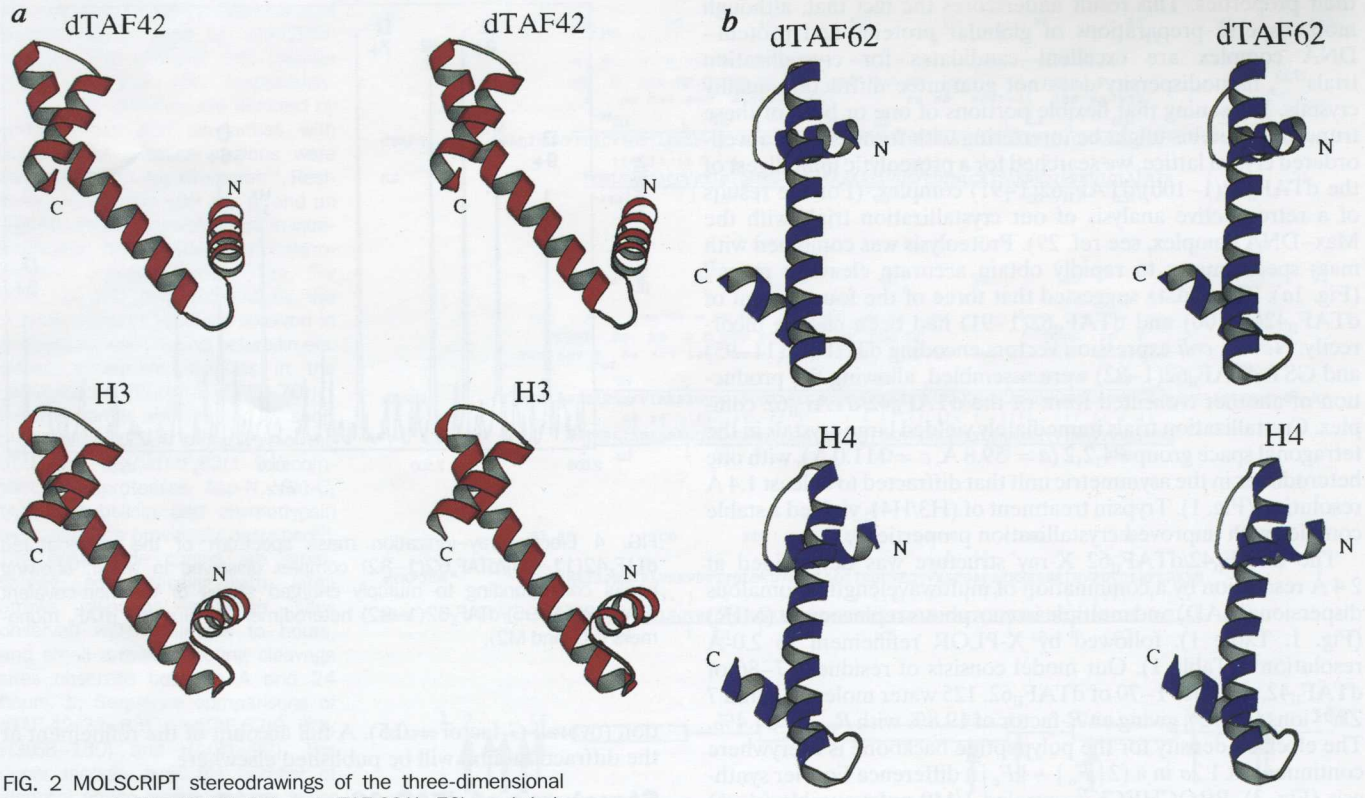


FIG. 2 MOLSCRIPT stereodrawings of the three-dimensional structures of $dTAF_{II}42(17–86)$ and $dTAF_{II}62(1–70)$, and their binary complex $dTAF_{II}42(17–86)/dTAF_{II}62(1–70)$. The corresponding views of histones H3 and H4, provided by G. Arents and E. N. Moudrianakis, have been included for comparison. Energetically favourable interactions include salt bridges (<4 Å), hydrogen bonds (<3.4 Å), van der Waals contacts (<4 Å) and two water-mediated bridges. *a*, $dTAF_{II}42(17–86)$ (top) and H3 (bottom). The additional N-terminal helix of H3 visualized in the structure of the histone octamer core has been omitted for clarity. *b*, $dTAF_{II}62(1–70)$ (top) and H4 (bottom). *c*, $dTAF_{II}42(17–86)/dTAF_{II}62(1–70)$ (top) and H3/H4 (bottom). There is one water molecule trapped between the two long α-helices (see Fig. 3).

Moreover, HNF3-α, a related factor, stabilizes a precisely positioned nucleosomal array in the enhancer of the mouse albumin gene, where it may function as a sequence-specific linker histone[26].

Here we report the crystal structure of a complex of two RNA polymerase II-specific, TAFs from *Drosophila*. The 2.0-Å resolution structure reveals that the N-terminal portions of $dTAF_{II}42$ and $dTAF_{II}62$ are folded into the canonical histone motif. Like histones H3 and H4, $dTAF_{II}42$ and $dTAF_{II}62$ form an intimate heterodimer. Within the crystal lattice two-fold symmetry generates a $dTAF_{II}42/dTAF_{II}62$ dimer of heterodimers, which exists in solution and resembles the histone $(H3/H4)_2$ heterotetramer. We have provided what we believe to be the first high-resolution view of the histone fold, and demonstrated another structural connection between the macromolecular machines responsible for transcription and DNA packaging.

Crystallization and structure determination

Following on from the first delineation of histone-like portions of $dTAF_{II}42$ and $dTAF_{II}62$ from primary sequence alignments[16], we expressed $dTAF_{II}42(1–100)$ and $dTAF_{II}62(1–91)$ separately as inclusion bodies in *Escherichia coli*, with the latter as a glutathione *S*-transferase (GST) fusion protein. Equimolar amounts of denatured $dTAF_{II}42(1–100)$ and $dTAF_{II}62(1–91)$ were co-renatured and purified to homogeneity. The resulting $dTAF_{II}42(1–100)/dTAF_{II}62(1–91)$ complex preparation was monodisperse and tetrameric, as judged by dynamic light scattering (data not shown), and largely α-helical, as judged by circular dichroism spectroscopy (data not shown).

Initial crystallization trials immediately yielded small crystals, which diffracted weakly to 4–5 Å and resisted attempts to improve

their properties. This result underscores the fact that, although monodisperse preparations of globular proteins and protein–DNA complex are excellent candidates for crystallization trials[27,28], monodispersity does not guarantee diffraction-quality crystals. Reasoning that flexible portions of one or both of these truncated proteins might be interfering with formation of a well-ordered crystal lattice, we searched for a proteolytic limit digest of the $dTAF_{II}42(1–100)/dTAF_{II}62(1–91)$ complex. (For the results of a retrospective analysis of our crystallization trials with the Max–DNA complex, see ref. 29). Proteolysis was combined with mass spectrometry to rapidly obtain accurate cleavage maps[29] (Fig. 1*a*). These data suggested that three of the four termini of $dTAF_{II}42(1–100)$ and $dTAF_{II}62(1–91)$ had been chosen incorrectly. New *E. coli* expression vectors encoding $dTAF_{II}42(11–95)$ and GST-$dTAF_{II}62(1–82)$ were assembled, allowing the production of another truncated form of the $dTAF_{II}42/dTAF_{II}62$ complex. Crystallization trials immediately yielded large crystals in the tetragonal space group $P4_12_12$ ($a = 59.8$ Å, $c = 111.0$ Å), with one heterodimer in the asymmetric unit that diffracted to at least 1.4 Å resolution (Fig. 1). Trypsin treatment of $(H3/H4)_2$ yielded a stable complex with improved crystallization properties[30].

The $dTAF_{II}42/dTAF_{II}62$ X-ray structure was determined at 2.4 Å resolution by a combination of multiwavelength anomalous dispersion (MAD) and multiple isomorphous replacement (MIR) (Fig. 1, Table 1), followed by X-PLOR refinement to 2.0-Å resolution[31] (Table 1). Our model consists of residues 17–86 of $dTAF_{II}42$, residues 1–70 of $dTAF_{II}62$, 125 water molecules, and 7 Zn^{2+} ions (Fig. 2), giving an *R*-factor of 19.8% with $R_{free} = 24.4\%$. The electron density for the polypeptide backbone is everywhere continuous at 1.2σ in a $(2|F_o| - |F_c|)$ difference Fourier synthesis (Fig. 3). PROCHECK[32] revealed 1/140 unfavourable (ϕ, ψ) combinations, and main-chain and side-chain structural parameters consistently better than those expected at 2.0-Å resolution (overall *G*-factor = 0.5). A full account of the refinement at the diffraction limit will be published elsewhere.

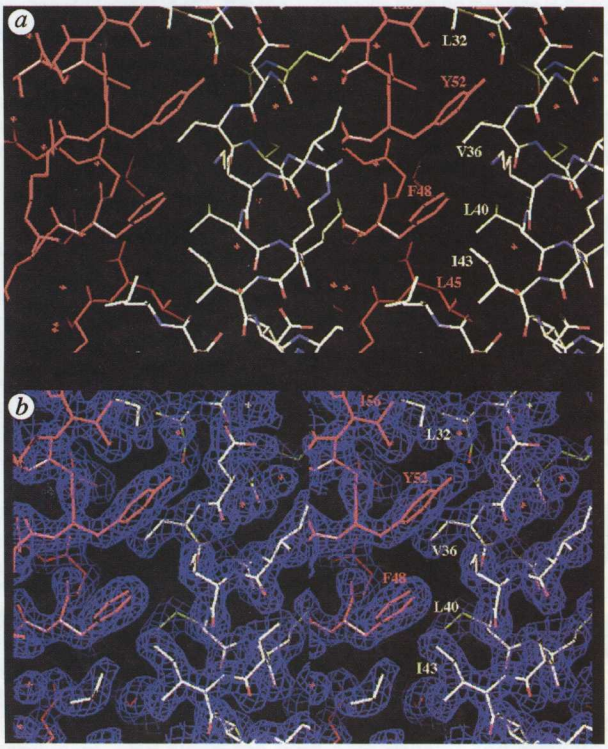


FIG. 3 *a*, Stereo drawing of the heterodimer interface, showing the interactions between $dTAF_{II}42(17–86)$ and $dTAF_{II}62(1–70)$. Critical hydrophobic residues at the interface are labelled with single-letter amino-acid code. The water molecule bridging Tyr 52 to *Asp 35* and Thr 81 to *Asp 35* can be seen above the OH group of Tyr 52. *b*, Stereo drawing of the 2.0-Å resolution $(2|F_o| - |F_c|)$ difference Fourier synthesis for the portion of the TAF_{II} complex depicted in *a*.

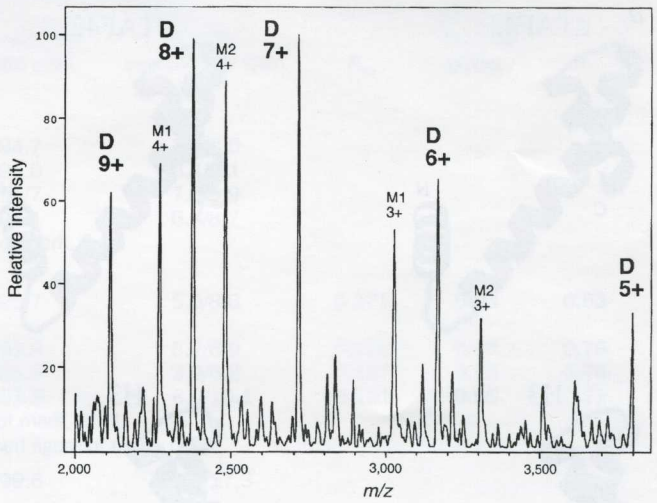


FIG. 4 Electrospray–ionization mass spectrum of the co-renatured $dTAF_{II}42(11–95)/dTAF_{II}62(1–82)$ complex dissolved in water, showing peaks corresponding to multiply charged states of the non-covalent $dTAF_{II}42(11–95)/dTAF_{II}62(1–82)$ heterodimer (D) plus the $dTAF_{II}$ monomers (M1 and M2).

Structures of $dTAF_{II}42$ and $dTAF_{II}62$

The X-ray structures of the histone-like regions of $dTAF_{II}42$ and $dTAF_{II}62$ are illustrated in Fig. 2. (For clarity, the TAF_{II} fragments used in this study are denoted $dTAF_{II}42(11–95)$ and $dTAF_{II}62(1–82)$. The portions of these two polypeptide chains visualized in the final electron-density map are denoted $dTAF_{II}42(17–86)$ and $dTAF_{II}62(1–70)$, with their histone-like regions denoted $dTAF_{II}42(22–83)$ and $dTAF_{II}62(9–70)$, respectively.) As predicted[16,19], the N-terminal portions of both $dTAF_{II}42$ and $dTAF_{II}62$ adopt the canonical histone motif[33], consisting of a long central α-helix flanked on each side by a short α-helix (Fig. 2). The truncated form of $dTAF_{II}42$ used for crystallization does not include the portion of the polypeptide chain corresponding to the additional H3 N-terminal α-helix, observed in the histone octamer structure[34]. Throughout this paper, we will describe structural details using α1, L1, α2, L2 and α3 to denote the segments within the histone fold (italics/sloping Greek denote $dTAF_{II}62$). Comparisons of the histone-homology regions of the TAF_{II} structures with histones H3 and H4 revealed the following root-mean-square deviations (r.m.s.ds) between α-carbon atomic positions: $dTAF_{II}42(22–83)$ versus $dTAF_{II}62(9–70)$, 1.6 Å; $dTAF_{II}42(22–83)$ versus H3(68–130), 1.6 Å; and $dTAF_{II}62(9–70)$ versus H4(31–92), 1.6 Å, respectively. These values compare favourably with those obtained by comparing H2A, H2B, H3 and H4 with one another (see Table 1 in ref. 33). Presumably these relatively large r.m.s.ds reflect differences in the trajectory of α2. In H3 the α2 helix is linear, but is bent near its C terminus in $dTAF_{II}42$; the converse is true for H4 and $dTAF_{II}62$.

The canonical histone fold has been postulated to arise from gene duplication[33], giving two helix–strand–helix (HSH) segments (HSH1 and HSH2 refer to the N- and C-terminal halves of the histone fold, respectively). The results of a detailed comparison of HSH1 from $dTAF_{II}42(22–51)$, $dTAF_{II}62(9–38)$, H3(68–98) and H4(31–60), and HSH2 from $dTAF_{II}42(52–83)$, $dTAF_{II}62(39–70)$, H3(99–130) and H4(61–92), are shown in Fig. 1. The HSH2 fold is very similar, displaying r.m.s.ds between α-carbon atomic positions of 0.7–1.0 Å. In contrast, the HSH1 fold is less similar, with the corresponding r.m.s.ds ranging from 1.4 to 1.9 Å. This marked difference between HSH1 and HSH2 may reflect the critical role of HSH2 in mediating interactions

between heterodimers within the histone octamer crystal structure[34] or between TAF_{II} heterodimers and homodimers.

Differential scanning calorimetry demonstrated a single cooperative unfolding transition for the H3/H4 complex[35]. By analogy, the histone-like portions of $dTAF_{II}42$ and $dTAF_{II}62$ are probably unable to assume their full native folds in the absence of one another, because they too demonstrate only a small number of long-range intramolecular contacts. Residues involved in these contacts have been referred to as 'self' residues[33] (denoted with 's' in Fig. 1*a*). For $dTAF_{II}42(17–86)$ these intramolecular interactions include α1–α2, α1–L1, α2–L1, α2–L2, α2–α3 and L2–α3; for $dTAF_{II}62(1–70)$, they include *α1–α2*, *α1–L1*, *α2–L1*, *α2–L2*, *α2–α3* and *L2–α3* (italics/sloping Greek are used throughout this paper to refer to $dTAF_{II}62$).

Structure of $dTAF_{II}42/dTAF_{II}62$

The structure of the $dTAF_{II}42(17–86)/dTAF_{II}62(1–70)$ heterodimer is shown in Fig. 2. As in the H3/H4 heterodimer, also illustrated in Fig. 2, the two histone folds interact with one another in a head-to-tail fashion[33]. Stabilizing contacts between $dTAF_{II}42(17–86)$ and $dTAF_{II}62(1–70)$ are largely hydrophobic, span the entire length of both molecules, and are conserved with H3 and H4 (Fig. 1). Binary complex formation buries about 3,390 Å^2 of solvent-accessible surface area (56% of the buried surface is hydrophobic, with the remainder either polar or charged). The residues involved in heterodimer contacts have been referred to as 'pair' residues[33] (denoted with 'p' in Fig. 1*a*). $dTAF_{II}42(17–86)/dTAF_{II}62(1–70)$ heterodimer interactions involve each segment of the two polypeptide chains, including α1–*α1*, α1–α2, α1–*L2*, L1–α2, L1–*L2*, L1–*α3*, α2–*α3*, α2–α2, α1–*L1*, α2–*α1*, L2–*L1*, L2–*α2* and α3–α2. Asymmetry in these interactions probably explains why H3, H4, $dTAF_{II}42$ and $dTAF_{II}62$ do not show any significant propensity to form homodimers. Electrospray ionization mass spectrometry was also used to examine the stability of the $dTAF_{II}42(11–95)/dTAF_{II}62(1–82)$ heterodimer. Under moderately gentle electrospray conditions, a pure water solution of co-renatured $dTAF_{II}42(11–95)/dTAF_{II}62(1–82)$ exhibits strong multiply charged peaks corresponding to the non-covalent heterodimer and the individual monomers (Fig. 4).

Structure of $(dTAF_{II}42/dTAF_{II}62)_2$

The structure of the $(dTAF_{II}42(17–86)/dTAF_{II}62(1–70))_2$ heterotetramer is shown in Fig. 5. As in the histone octamer[34], the symmetry axis within the TAF_{II} tetramer coincides with a crystallographic two-fold. Interactions involving the HSH2 portion of the H3 homologue, $dTAF_{II}42(17–86)$, stabilize the heterotetramer, burying about 670 Å^2 of solvent-accessible surface area (48% of the buried surface is hydrophobic, with the remainder either polar or charged). These values are typical for protein–protein complexes (reviewed in ref. 36), and are consistent with the equilibrium dissociation constant of 10^{-6} M (Fig. 1). Residues involved in $dTAF_{II}42(17–86)$–$dTAF_{II}42(17–86)$ interactions in the crystalline state are denoted with an asterisk in Fig. 1*a*, and correspond to conversed residues in H3.

We presume that the same residues (Asp 59, Val 62, Tyr 63, His 66, Leu 79, Glu 82 and Val 83) stabilize the solution form of the $(dTAF_{II}42/dTAF_{II}62)_2$ heterotetramer, but cannot be certain without experimental confirmation. The proteolysis–mass spectrometry results provide indirect support for this prediction. Enzymatic cleavage at three sites (Tyr 63, Arg 68 and Asp 76) within the HSH2 region of $dTAF_{II}42(1–100)$ is slow compared with cleavage events outside the histone-homology region (Fig. 1*a*). Two of these three sites are solvent inaccessible in the crystallographic heterotetramer, and all three are completely exposed in the crystallographic heterodimer.

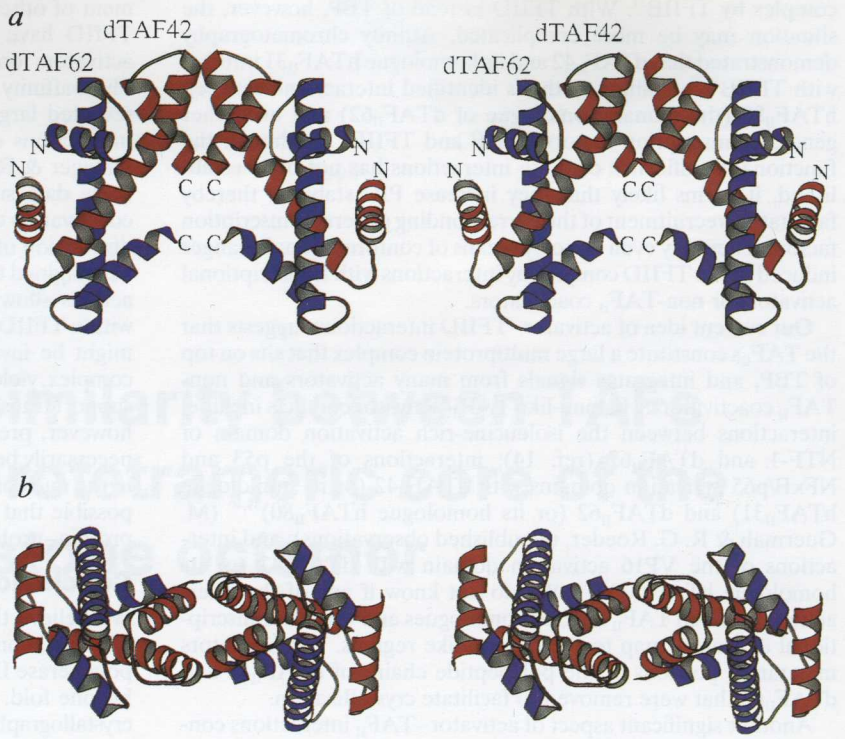


FIG. 5 Stereo drawing ribbon representations of the $(dTAF_{II}42(17–86)/dTAF_{II}62(1–70))_2$ heterotetramer, generated by two-fold crystallographic symmetry. *a*, View perpendicular to the two-fold symmetry axis. *b*, View along the two-fold symmetry axis.

TFIID assembly

Our biophysical study of the $dTAF_{II}42/dTAF_{II}62$ complex suggests that TFIID contains a $(dTAF_{II}42/dTAF_{II}62)_2$ heterotetramer, which could interact with H2B-like TAF_{II}s to form a histone-like octamer. Compelling, albeit indirect, support for these predictions comes from the results of recent studies of the human TAF_{II} homologue of histone H2B ($hTAF_{II}20$)[18]. The histone-like region of $hTAF_{II}20$ alone allows assembly of TFIID *in vivo*, the measured $hTAF_{II}20$:TBP ratio in TFIID is 4 : 1, and an octamer-like pattern of protein–protein interactions has been documented for $hTAF_{II}31$ (homologous to H3 and $dTAF_{II}42$), $hTAF_{II}80$ (homologous to H4 and $dTAF_{II}62$), and $hTAF_{II}20$ (homologous to H2B and $dTAF_{II}30\alpha$)[18]. In addition, the histone-like regions of $dTAF_{II}62$ and $dTAF_{II}30\alpha$ contain HSH2 residues homologous, respectively, to the H4 and H2B residues responsible for mediating H4–H2B interactions in the histone octamer structure[34] (Fig. 1*a*). Protein–protein interactions have also been demonstrated between $hTAF_{II}80$ and TBP, and between $hTAF_{II}20$ and TBP, $hTAF_{II}135$ and $hTAF_{II}55$ (ref. 18). Thus TFIID may contain a histone-like octameric TAF_{II} substructure $((hTAF_{II}20)_2–(hTAF_{II}31/dTAF_{II}80)_2–(hTAF_{II}20)_2$ or $(dTAF_{II}30\alpha)_2–(dTAF_{II}42/dTAF_{II}62)_2–(dTAF_{II}30\alpha)_2)$, anchored to TBP, $hTAF_{II}135$ (or $dTAF_{II}110$) and $hTAF_{II}55$ (there is no known *Drosophila* homologue for this human TAF_{II}).[18] There are other TBP–TAF_{II} and TAF_{II}–TAF_{II} interactions (reviewed in ref. 9) that probably stabilize the remainder of the TFIID complex.

The co-crystal structure of core TFIIB recognizing the preformed TBP–DNA complex showed the second step of PIC assembly, explaining stabilization of the TBP–TATA element

complex by TFIIB[37]. With TFIID instead of TBP, however, the situation may be more complicated. Affinity chromatography demonstrated that $dTAF_{II}42$ and its homologue $hTAF_{II}31$ interact with TFIIB[20,38]. Similar methods identified interactions between $hTAF_{II}80$ (the human homologue of $dTAF_{II}62$) and two other general transcription factors TFIIF and TFIIE[17]. Although the functional significance of these interactions has not been established, it seems likely that they increase PIC stability, thereby facilitating recruitment of the corresponding general transcription factors. They may even do so by means of conformational changes induced in the TFIID complex by interactions with transcriptional activators or non-TAF_{II} coactivators.

Our current idea of activator–TFIID interactions suggests that the TAF_{II}s constitute a large multiprotein complex that sits on top of TBP, and integrates signals from many activators and non-TAF_{II} coactivators. Histone-like TAF_{II}-activator contacts include: interactions between the isoleucine-rich activation domain of NTF-1 and $dTAF_{II}62$ (ref. 14); interactions of the p53 and NFκB/p65 activation domains with $dTAF_{II}42$ (or its homologue $hTAF_{II}31$) and $dTAF_{II}62$ (or its homologue $hTAF_{II}80$)[39,40] (M. Guermah & R. G. Roeder, unpublished observations); and interactions of the VP16 activation domain with $dTAF_{II}42$ (or its homologue $hTAF_{II}31$)[20,38]. We do not know if any of the interactions between TAF_{II} histone homologues and various transcriptional activators map to the histone-like regions. The activators may target portions of the polypeptide chains of $dTAF_{II}42$ and $dTAF_{II}62$ that were removed to facilitate crystallization.

Another significant aspect of activator–TAF_{II} interactions concerns induced conformational changes in TFIID and TFIID-promoter complexes. Activator-induced changes have been demonstrated in TFIID-promoter complex[41,42] manifested by downstream extension of the TFIID footprint well beyond the transcription start site that was correlated with increased recruitment of other general factors. Later studies with highly purified TFIID have confirmed qualitative and quantitative effects of activators on TFIID binding, sometimes requiring TFIIA[43,44]. Photoaffinity DNA–protein crosslinking studies have also revealed large changes in TAF_{II}–DNA contacts in response to interactions of TFIIA with promoter-bound TFIID (T. Oelgeschlager & R. G. Roeder, unpublished observations). Together, these data suggest that binding of transcriptional activators or coactivators can cause substantial rearrangements in the relative disposition of TFIID subunits and DNA. In this context, it might be imagined that activator-induced changes in TAF_{II}–DNA interactions allow the presumptive histone octamer-like substructure within TFIID to engage DNA. We speculate that such a structure might be involved in stabilizing an activator–TFIID–promoter complex, yielding a stereospecific nucleoprotein assembly that can support transcriptional activation (reviewed in ref. 45). We do not, however, predict that this putative octamer substructure would necessarily be encircled by a piece of DNA 146 base pairs long, as in the nucleosome core particle (reviewed in refs 46–49). It is also possible that TFIID merely exploits the histone fold to mediate protein–protein contacts between various TAF_{II}s.

Conclusion

We believe that the 2.0-Å resolution structure of the $dTAF_{II}42$/$dTAF_{II}62$ complex provides the first structural knowledge of RNA polymerase II TAFs, and the first high-resolution structure of the histone fold. Our work also provides a starting point for further crystallographic, biochemical and genetic studies of TFIID assembly and its role in Pol II-mediated transcription activation. Finally, studies of the three known histone-like TAF_{II}s suggest an evolutionary relationship between the histone octamer and a TFIID substructure that may be functionally significant. □

Received 13 December 1995; accepted 9 February 1996.

1. Roeder, R. G. *Trends biochem. Sci* **16,** 402–408 (1991).
2. Hori, R. & Carey, M. *Curr. Opin. Genet. Dev.* **4,** 236–244 (1994).
3. Zawel, L. & Reinberg, D. *Prog. Nucleic Acid Res. molec. Biol.* **44,** 67–108 (1993).
4. Zawel, L., Kumar, K. & Reinberg, D. *Genes Dev.* **9,** 1479–1490 (1995).
5. Koleske, A. & Young, R. *Trends biochem. Sci.* **20,** 113–116 (1995).
6. Matsui, T., Segall, J., Weil, P. & Roeder, R. G. *J. biol. Chem.* **255,** 11992–11996 (1980).
7. Sawadogo, M. & Roeder, R. G. *Cell* **43,** 165–175 (1985).
8. Kim, J. L. & Burley, S. K. *Nature Struct. Biol.* **1,** 638–653 (1994).
9. Burley, S. K. & Roeder, R. G. *A. Rev. Biochem.* **65** (in the press).
10. Owen-Hughes, T. & Workman, J. L. *Crit. Rev. Euk. Gene Express.* **4,** 403–441 (1994).
11. Nikolov, D. B. & Burley, S. K. *Nature Struct. Biol.* **1,** 621–637 (1994).
12. Reese, J., Apone, L., Walker, S., Griffin, L. & Green, M. *Nature* **371,** 523–527 (1994).
13. Poon, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **92,** 8224–8228 (1995).
14. Chen, J., Attardi, L., Verrijzer, D., Yokomori, K. & Tjian, R. *Cell* **79,** 93–105 (1994).
15. Sauer, F., Hansen, S. & Tjian, R. *Science* **270,** 1783–1788 (1995).
16. Kokubo, T. *et al. Nature* **367,** 484–487 (1994).
17. Hisatake, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **92,** 8195–8199 (1995).
18. Hoffmann, A. *et al. Nature* **380,** 356–359 (1996).
19. Baxevanis, A., Arents, G., Moudrianakis, E. & Landsman, D. *Nucleic Acids Res.* **23,** 2685–2691 (1995).
20. Goodrich, J., Hoey, T., Thut, C., Admon, A. & Tjian, R. *Cell* **75,** 519–30 (1993).
21. Weinzierl, R., Ruppert, S., Dynlacht, B., Tanese, N. & Tjian, R. *EMBO J.* **12,** 5303–5309 (1993).
22. Mengus, G. *et al. EMBO J.* **14,** 1520–1531 (1995).
23. Yokomori, K., Chen, J.-L., Admon, A., Zhou, S. & Tjian, R. *Genes Dev.* **7,** 2587–2597 (1993).
24. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. *Nature* **364,** 412–420 (1993).
25. Ramakrishnan, V., Finch, J., Graziano, V. & Sweet, R. *Nature* **362,** 219–223 (1993).
26. McPherson, C., Shin, E.-Y., Friedman, D. & Zaret, K. *Cell* **75,** 387–398 (1993).
27. D'Arcy, A. *Acta crystallogr.* **D50,** 469–411 (1994).
28. Ferré-D'Amaré, A. R. & Burley, S. K. *Structure* **2,** 357–359, 567 (1994).
29. Cohen, S. L., Ferre-D'Amare, A. R., Burley, S. K. & Chait, B. T. *Protein Sci.* **4,** 1088–1099 (1995).
30. Lattman, E., Burlingame, R., Hatch, C. & Moudrianakis, E. N. *Science* **216,** 1016–1018 (1982).
31. Brünger, A. T. *X-PLOR, Version 3.1 manual* (Yale University, New Haven, CT, 1992).
32. Laskowski, R. J., Macarthur, M. W., Moss, D. S. & Thornton, J. M. *J. appl. Crystallogr.* **26,** 283–290 (1993).
33. Arents, G. & Moudrianakis, E. *Proc. natn. Acad. Sci. U.S.A.* **92,** 11170–11174 (1995).
34. Arents, G., Burlingame, R. W., Wang, B.-C., Love, W. E. & Moudrianakis, E. N. *Proc. natn. Acad. Sci. U.S.A.* **88,** 10148–10152 (1991).
35. Karantza, V., Friere, E. & Moudrianakis, E. N. *Biochemistry* **35** (in the press).
36. Janin, J. *Proteins* **21,** 30–39 (1995).
37. Nikolov, D. B. *et al. Nature* **377,** 119–128 (1995).
38. Klemm, R., Goodrich, J., Zhou, S. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **92,** 92, 5788–5792 (1995).
39. Thut, C., Chen, J. L., Klemm, R. & Tjian, R. *Science* **267,** 100–104 (1995).
40. Lu, H. & Levine, A. *Proc. natn. Acad. Sci. U.S.A.* **92,** 5154–5158 (1995).
41. Horikoshi, M., Carey, M., Kadidani, H. & Roeder, R. G. *Cell* **54,** 665–669 (1988).
42. Horikoshi, M., Hai, T., Lin, Y.-S., Green, M. R. & Roeder, R. G. *Cell* **54,** 1033–1042 (1988).
43. Lieberman, P. & Berk, A. *Genes Dev.* **8,** 995–1006 (1994).
44. Chi, T., Lieberman, P., Ellwood, K. & Carey, M. *Nature* **377,** 254–257 (1995).
45. Tjian, R. & Maniatis, T. *Cell* **77,** 5–8 (1994).
46. Klug, A., Rhodes, D., Smith, J., Finch, J. & Thomas, J. *Nature* **287,** 509–516 (1980).
47. Richmond, T., Finch, J., Rushton, B., Rhodes, D. & Klug, A. *Nature* **311,** 532–537 (1984).
48. Arents, G. & Moudrianakis, E. N. *Proc. natn. Acad. Sci. U.S.A.* **90,** 10489–10493 (1993).
49. Pruss, D., Hayes, J. & Wolffe, A. *BioEssays* **17,** 161–170 (1995).
50. Hendrickson, W. A. *Science* **254,** 51–58 (1991).

ACKNOWLEDGEMENTS. We thank G. Arents and E. N. Moudrianakis for access to their work before publication, help with figures, and discussion; C. Ogata for help using the X4A Beamline; M. Capel for help using the X12B Beamline; S. Ramakoti for assistance with X-ray measurements; A. Ferré-D'Amaré, J. Goldberg, H. Houbaviy, J. L. Kim, J. Kuriyan, D. B. Nikolov, G. Patikoglou and G. A. Petsko for discussions; and A. Gazes and the staff of Rockefeller University Computing Services for technical support. This work was supported in part by the Howard Hughes Medical Institute (S.K.B.), the NIH (R.G.R., B.C.T.) and the Human Frontiers Scientific Program (R.G.R.).