

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, TFIIIF, TFIIG/J, TFIIH and TFIIH (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 multi-protein-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 TFIIIF. *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 dTAF₃₀ α , dTAF₄₂ and dTAF₆₂ 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 structures³⁴. Residues are labelled with 's', 'p' and an asterisk, indicating involvement in intramolecular, heterodimer, and heterotetramer contacts, respectively. For H2B, H4, dTAF₃₀ α and dTAF₆₂, the asterisk denotes residues involved in contacts between distinct heterodimers. Solvent-accessible residues in the (dTAF₄₂(17–86))/(dTAF₆₂(1–70))₂ heterotetramer are denoted by short black bars. Bottom, proteolysis of the dTAF₄₂(1–100)/dTAF₆₂(1–91) complex. Endoproteases Asp-N, Glu-C, trypsin, subtilisin and chymotrypsin were used as previously described²⁹.

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 dTAF₄₂(22–83), dTAF₆₂(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 dTAF₄₂s 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 dTAF₄₂(11–95)/dTAF₆₂(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 (dTAF₄₂(11–95)/dTAF₆₂(1–82))₂ heterotetramer to the dTAF₄₂(11–95)/dTAF₆₂(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 reaction³⁵.

METHODS. *Drosophila* TAF₄₂ residues 11–95 and *Drosophila* TAF₆₂ residues 1–82 fused with GST were separately overexpressed as inclusion bodies in *E. coli*. Denatured dTAF₄₂(11–95) and GST-dTAF₆₂(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 dTAF₄₂(11–95)/dTAF₆₂(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: dTAF₄₂(11–95), 9,917 \pm 1 (predicted, 9,918); and dTAF₆₂(1–82), 9,071 \pm 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.

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 discovery⁶. 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

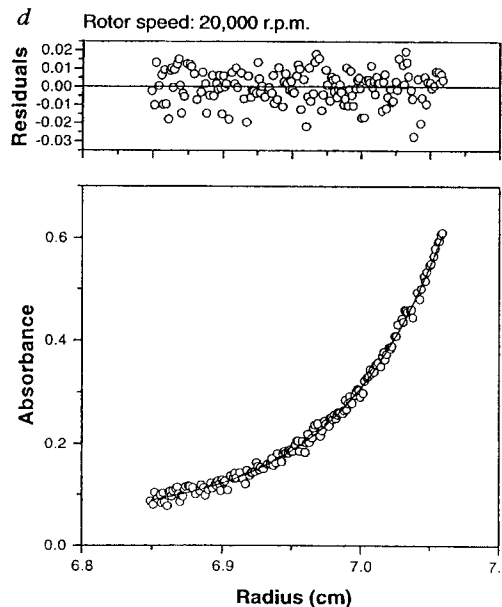
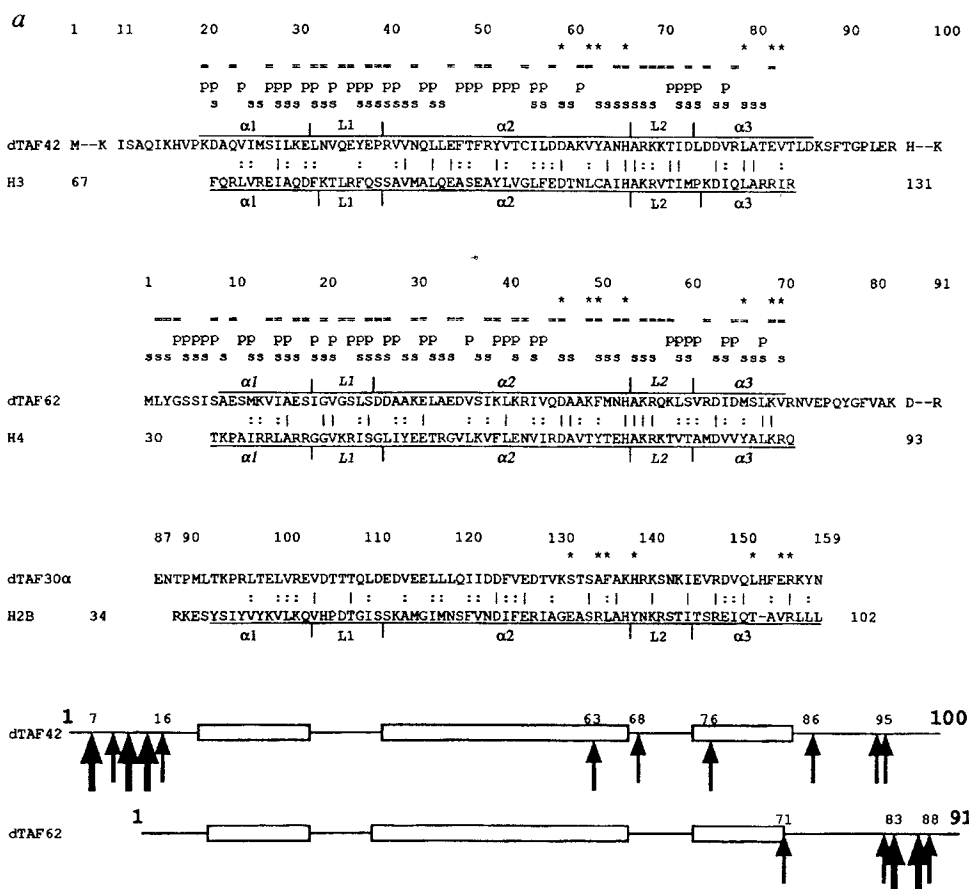


TABLE 1 Statistics of the crystallographic analysis

	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)							
$\lambda 1$ (0.9873 Å)	18.0–2.4	34,011/7,697	91.0/94.7	5.9/8.0			
$\lambda 2$ (0.9795 Å)	18.0–2.4	33,699/7,643	91.2/95.0	6.7/9.1			
$\lambda 3$ (0.9792 Å)	18.0–2.4	33,786/7,623	91.1/94.7	7.6/9.9			
$\lambda 4$ (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 forms⁵⁰, 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}62(1–82) complex against HgCl₂. Oscillation photography at 100K and data reduction were performed as previously described⁸. 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 described¹⁴. MAD⁵⁰ 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 Å², the 125 water molecules included in the model have an average *B*-factor of 29 Å², and the 7 Zn²⁺ ions have an average *B*-factor of 28 Å². Atomic coordinates and structure factor amplitudes will be submitted to the Brookhaven Protein Data Bank. $R_{sym} = \sum |I - \langle I \rangle| / \sum I$, 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 = $\sum ||F_{PH}| - |F_P|| / \sum |F_P|$, 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. ($|F_{PH}| / E$), where $|F_{PH}|$ 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 archaeobacterium (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 recruitment¹⁵.

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}30 α and hTAF_{II}20)^{16,18,22,23} that are putative histone H2B homologues¹⁸, 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 DNA²⁵.

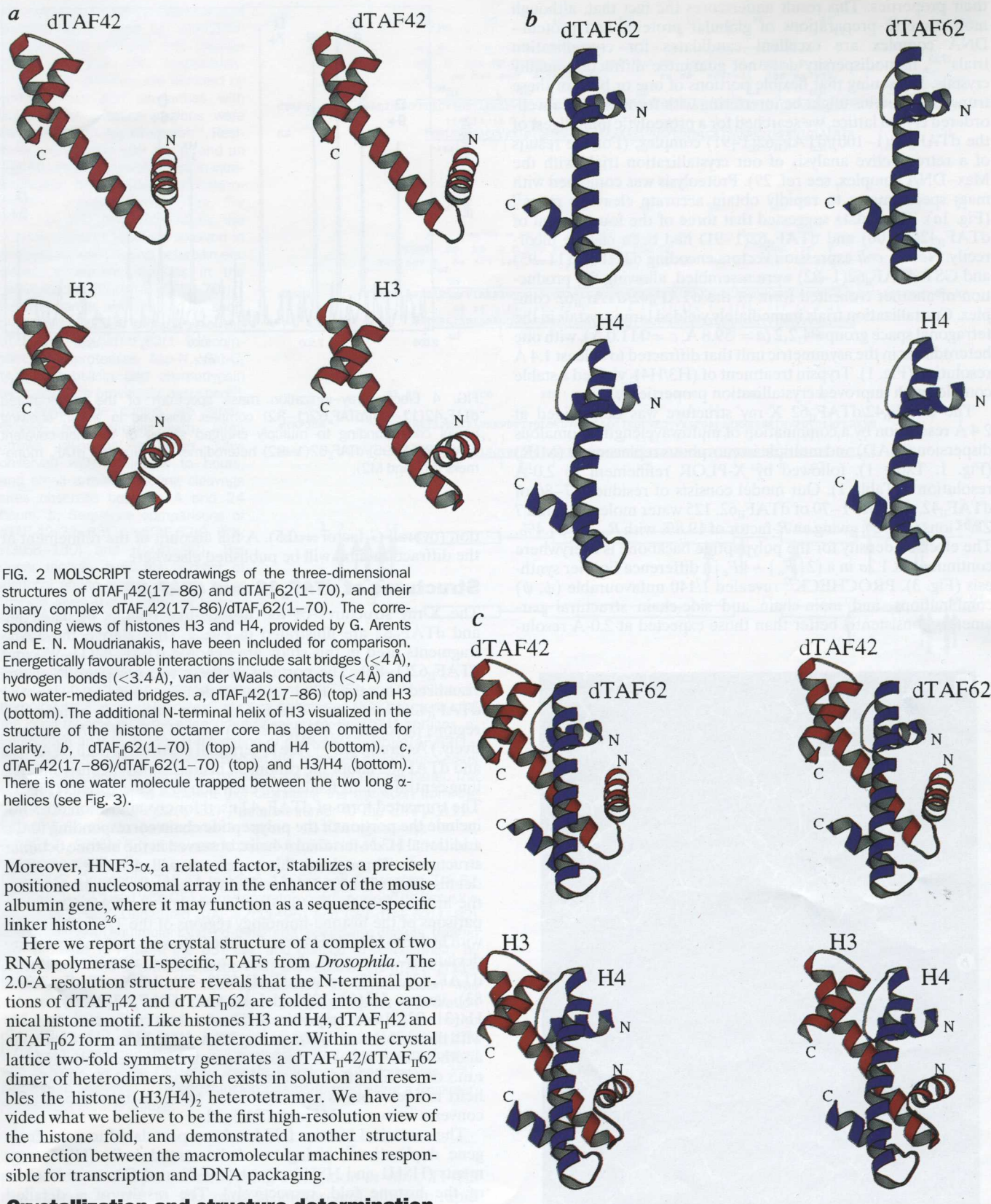


FIG. 2 MOLSCRIPT stereodrawings of the three-dimensional structures of dTAF_{II}42(17–86) and dTAF_{II}62(1–70), and their binary complex dTAF_{II}42(17–86)/dTAF_{II}62(1–70). The corresponding views of histones H3 and H4, provided by G. Arents and E. N. Moudrianakis, have been included for comparison. 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 histone²⁶.

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 alignments¹⁶, 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 maps²⁹ (Fig. 1a). 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 *P*4₂,2 (*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 properties³⁰.

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 resolution³¹ (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 Zn²⁺ 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). PROCHECK³² revealed 1/140 unfavourable (*φ*, *ψ*) combinations, and main-chain and side-chain structural parameters consistently better than those expected at 2.0-Å resolu-

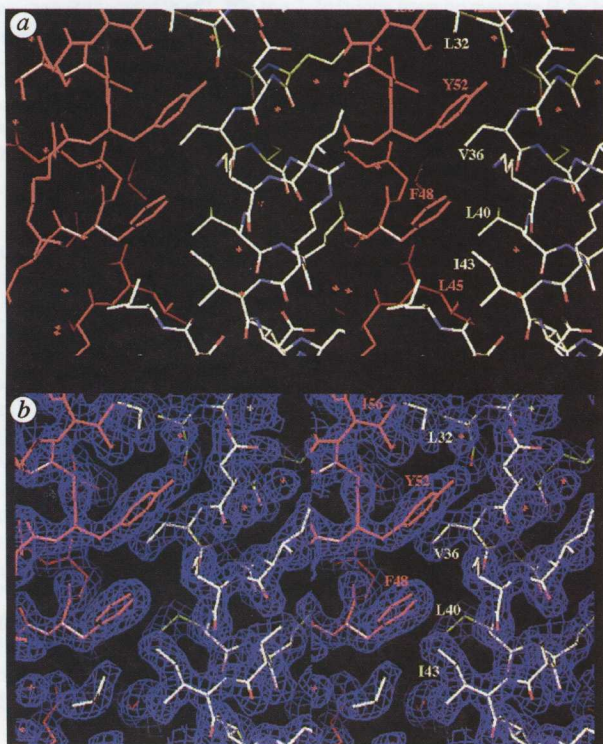


FIG. 3 a, Stereo drawing of the heterodimer interface, showing the interactions between dTAF_{II}42(17-86) and dTAF_{II}62(1-70). Critical hydrophobic residues at the interface are labelled with single-letter amino-acid code. The water molecule bridging Tyr 52 to Asp 35 and Thr 81 to Asp 35 can be seen 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.

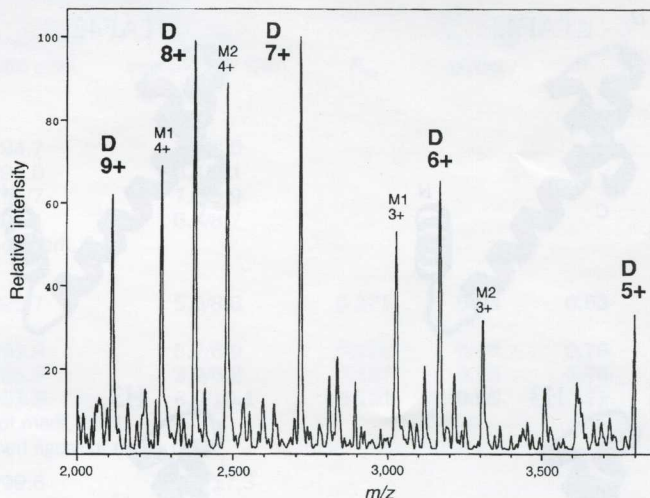


FIG. 4 Electrospray-ionization mass spectrum of the co-renatured dTAF_{II}42(11-95)/dTAF_{II}62(1-82) complex dissolved in water, showing peaks corresponding to multiply charged states of the non-covalent dTAF_{II}42(11-95)/dTAF_{II}62(1-82) heterodimer (D) plus the dTAF_{II} monomers (M1 and M2).

tion (overall *G*-factor = 0.5). A full account of the refinement at the diffraction limit will be published elsewhere.

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 motif³³, 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 structure³⁴. 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 duplication³³, 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

